A Review on Deep Reinforcement Learning-Based UAV Navigation and Conflict Resolution in Unknown Environments

Harshvardhan Pandey
Dept. of Artificial Intelligence and Machine Learning
Manipal University Jaipur
Jaipur, India
harshpandey145@gmail.com

Adya Chauhan
Dept. of Artificial Intelligence and Machine Learning
Manipal University Jaipur
Jaipur, India
adyachauhan.04@gmail.com

Yash Prasad
Dept. of Artificial Intelligence and Machine Learning
Manipal University Jaipur
Jaipur, India
eyash.prasad24@gmail.com

***Abstract*—Unmanned Aerial Vehicles or UAVs have become very popular. They have found applications in diverse fields including surveillance, disaster response, logistics and inspections of unsafe situations and defence missions. However, their ability to navigate through complex, dynamic environments in case of communication loss is important. Traditional approaches lack adaptability and are computationally expensive. In response to this, researchers have integrated Deep Reinforcement learning techniques. This paper reviews six studies on DRL-based and hybrid approaches for UAV navigation and collision avoidance. They are analyzed based on problem formulation, approaches, architectural innovations, performance and real-world applications. A comparative study is provided evaluating their strengths, limitations, and potential for future research.**

***Index Terms*—unmanned aerial vehicles, deep reinforcement learning, navigation, obstacle avoidance, collision resolution, autonomous systems**

I. Introduction

A. Background and Motivation

Unmanned aerial vehicles (UAVs), widely known as drones, have come to occupy a vital place in today's automation environment. Their compactness and easy manoeuvrability at an ever-decreasing cost have opened their applications into the domains of military, commercial, and industrial. One such huge challenge is the autonomous navigation of these vehicles in entirely unknown or partly known environments. Classical path generation techniques like A*, Rapidly exploring random trees (RRT), and artificial potential fields have the limitation of being dependent on pre-existing maps and are computationally inefficient in dynamic environments. They are designed and trained for static ideal-world scenarios, making them unreliable in unknown environments.

Deep reinforcement learning (DRL) is a promising paradigm that has transformed the approach to solving navigation problems. This approach allows agents to learn optimal policies through experiences accumulated in simulated environments, notwithstanding the criteria of partial observability and dynamic constraints. Thus, DRL has allowed navigation policy learning that signifies on selection of optimal policy based on prior training. Integrating high-dimensional inputs such as images and LIDAR data while accommodating real-time decision-making without exhaustive path computation makes these models ideal for dynamic environments.

B. Objectives and Scope

This review critically analyzes the following seven key papers:

1) Deep-reinforcement-learning-based UAV autonomous navigation and collision avoidance in unknown environments (FRDDM-DQN)
2) Adaptive Multi-Agent Reinforcement Learning Solver for Tactical Conflict Resolution
3) UAV Navigation in 3D Urban Environments with Curriculum-based Deep Reinforcement Learning
4) Combining Motion Planner and Deep Reinforcement Learning for UAV Navigation in Unknown Environment
5) Autonomous UAV Navigation with Adaptive Control Based on Deep Reinforcement Learning
6) Multi-Level-Frontier Empowered Adaptive Path Planning for UAVs in Unknown Environments

Each paper is analyzed with respect to problem formulation, model architecture, training methodology, evaluation metrics, and experimental outcomes. The structure of this paper is as follows: Section 2 reviews related works and methodologies, Section 3 highlights indepth analysis of each paper. Section 4 offers the comparative analysis along with a proposed methodology, while Section 5 concludes the review.

II. Literature Review

The integration of Deep Reinforcement Learning (DRL) in the field of Unmanned Aerial Vehicles has attracted significant attention particularly based on partially observable and dynamic environments. This section presents a few studies that contributed towards this field.

Wang et al. [1] proposed the FRDDM-DQN framework that integrated Faster R-CNN for visual obstacle detection in place of convolutional neural networks and a Deep Q-Network (DQN) for policy learning. They used a custom mechanism called Data Deposit Mechanism (DDM) which prioritized safety-critical experiences in experience replay, promoting convergence. Though the method achieves superior performance in the simulated environments, higher computational resource requirements and a lack of verification in real-world settings remain major challenges.

Fremond et al. [2] set out the Multi-Agent Reinforcement Learning (MARL) framework that utilized Proximal Policy Optimization (PPO) with Recurrent Neural Networks (RNNs) to address the tactical conflict resolution of urban sky spaces. They adhered to the centralized training with decentralized execution (CTDE) paradigm. This enforces scalability and robustness in different conflict scenarios of the domain. However, the presumption of perfect communication and full-state observability limits the applicability to real-world scenarios.

Carvalho et al. [3] implemented a curriculum learning methodology in DQN for facilitating UAV navigation within complex 3D cities. They revealed that while increasing the difficulty by stepwise additions improved the generalization of policy and speed of convergence, their work did not involve any temporal modelling components, like Long Short-Term Memory (LSTM) networks, thereby affecting performance when one talked of dynamic environments.

Xue and Chen [4] made RLPlanNav. They combined a top-level DRL agent with a constantly working motion planner on the minor end (EGO-planner), where the high-level plan is generated and used for trajectory generation. This high-level DRL agent architecture complements the good performance of learning-based decision-making. It made the low-end motion planner dependent on deterministic planning, making the outcomes smoother and more feasible; however, this experiment was conducted at static environment types, thus raising doubts on its applicability towards dynamic context scenarios with simple shape obstacles.

Yin et al. [5] introduced the GARTD3, a framework for DRL with attention mechanisms and velocity-constrained loss functions that realize adaptive control in 3-dimensional environments. They used LSTMs to provide temporal awareness and also incorporated an attention module for navigation-obstacle avoidance balancing. Even though it was very effective, there were serious practical limitations regarding computational demands and sensitivity to hyperparameter tuning.

Duan et al. [6] described a non-learning-based method for extremely efficient path planning in unknown environments that used a multi-level octree voxelization. Historical State Record Tree (HSRT), aimed at recovering dead ends, is utilized for enhancing exploration efficiency. This method, while computationally light, is heuristic in nature, so it lacks adaptability and the capability to learn as DRL-based solutions do. An overview of the studies demonstrates varying strategies employed to enhance UAV navigation through DRL and hybrid systems. Despite the enormous progress made in this field, challenges like computational efficiency and the true applicability of UAVs, along with the adaptability of these UAVs in a dynamic environment remain something to ponder. Future research should aim at addressing these gaps in the current literature, possibly through lightweight modelling techniques, testing in a real environment, or utilising innovative learning paradigms.

III. In-Depth Analysis of Selected Works

A. FRDDM-DQN for UAV Navigation and Obstacle Avoidance

This work is concerned with autonomous UAV navigation in GPS-denied environments in which traditional sensors, like radar, are compromised by electronic jamming or simply do not work anymore. The actual challenge here is to enable some form of obstacle avoidance using visual information.

1) Methodology: The authors introduce a hybrid architecture that includes:

- Fast R-CNN for visually detecting obstacles
- DQNs for learning policies
- Data Deposit Mechanism (DDM) for prioritizing experience replay

Training takes place in a 3D simulation environment using image input. Faster R-CNN and DQN are separately trained as part of a training procedure that is designed to reduce the complexity of merging the two.

2) Innovations:

- Use of DDM, which selectively retains safety-critical experiences, enhancing convergence.
- A two-stage training pipeline to reduce the times required for retraining for dynamic scenarios.

3) Results: Outperformed YOLO-based methods and variants of the DQN (Dueling DQN, DDQN), achieving improved navigation success and convergence times while avoiding obstacles.

4) Limitations:

- Entirely simulation-based: no real-world deployment.
- Heavy computational costs from object detection during inference.

B. Adaptive Multi-Agent Reinforcement Learning

1) Problem Statement: Targets tactical conflict resolution (TCR) of UAVs within a shared urban airspace. It is an environment of decentralized UTM (unmanned traffic management) with heavy UAV traffic.

TABLE I: Comparison of DRL-based UAV Navigation Approaches (Part 1)

Author	Action	State	Reward	Algorithm	Environment
Wang et al. (2024)	Discrete actions from DQN	Camera images + positional info	Destination reached, collisions, out-of-bound penalties	FRDDM-DQN (Faster R-CNN + DQN)	3D simulated visual environment with varied obstacle layouts
Fremond et al. (2024)	Speed and altitude adjustments	Multi-UAV positional states	Safety, efficiency, conflict avoidance	Multi-Agent PPO with RNN	Urban airspace simulator with 9 tactical conflict scenarios
Carvalho et al. (2023)	Discrete 3D movements	Position, battery level, obstacle maps	Efficiency, collisions, energy-based shaping	Curriculum-Guided Deep Q-Learning	Simulated urban 3D maps with increasing complexity
Xue and Chen (2024)	Local waypoint targets	RGB image, partial maps	Goal achievement, smooth path, collisions	RLPlanNav (LSTM-DRL + Classical Planner)	Random static obstacle simulations with motion planner
Yin et al. (2024)	Continuous control: velocity + altitude	3D coordinates, obstacle vectors	Distance, obstacle avoidance, velocity constraints	Guide Attention TD3 (GARTD3) with LSTM + Attention	Complex 3D simulation with adaptive low-altitude flight
Duan et al. (2024)	Dynamic waypoints via frontier voxels	Multi-resolution voxel map	Forward movement, efficiency, dead-end penalties	Multi-Level Frontier Planner	Complex unknown voxel spaces with octree resolution control

TABLE II: Comparison of DRL-based UAV Navigation Approaches (Part 2)

Author	Novelty Factor	Limitations	Evaluation Metrics
Wang et al. (2024)	Faster R-CNN optimized for UAV kinematics + Data Deposit Mechanism	Limited to simulated environment; high preprocessing cost	Reward, navigation success rate, collision count, episode length
Fremond et al. (2024)	Centralized training with RNN generalization + ACAS-based evaluation	Assumes reliable comms; lacks real-world data	Conflict resolution rate, deviation from flight plan, success under Monte Carlo cases
Carvalho et al. (2023)	Curriculum learning + parallel training for speed	Q-learning stability issues in large spaces	Reward trajectory, convergence speed, success across curriculum levels
Xue and Chen (2024)	Hierarchical DRL + classical EGO planner integration	Limited to static obstacles; no dynamic reactivity	Path smoothness, trajectory feasibility, success rate
Yin et al. (2024)	Guide attention + adaptive velocity-constrained loss	High computation due to attention and memory	Navigation success, reward, collision rate, velocity adherence
Duan et al. (2024)	Multi-resolution voxel modeling + HSRT backtracking	Heuristic-based; not learning-capable	Planning time, dead-end escape rate, route efficiency

2) Methodology:

- Multi-Agent Reinforcement Learning (MARL) framework based on PPO
- Centralized Training and Decentralized Execution (CTDE)
- Common policy architecture across the UAVs
- Recurrent Neural Networks (RNN) adopted to generalize the varying intruder scenarios

3) Innovations:

- Validation under Airborne Collision Avoidance Systems(ACAS) standards
- Realistic airspace configurations including shared routes with single/multiple conflict points
- Nine synthetic case studies simulating a variety of conflict scenarios

4) Results:

- Conflict resolution above 99.9% in multi-agent encounters.
- Robustness across scenarios with varying agent density and dynamics.

5) Limitations:

- Simulation-based without hardware-in-the-loop testing
- Assumes full state observability and perfect communication

C. Curriculum-based DRL in 3D Urban Environments

1) Problem Statement: To navigate a broad expanse of 3D urban constructs with concerns on navigation, obstacle avoidance, and energy-optimization.

2) Methodology:

- Curriculum Learning with DQN
- Multi-agent training done in parallel
- Customized reward function concerning energy consumption, time, and distance

3) Innovations:

- Stages of curriculum with increasing complexity
- Transfer of policies effectively to an environment with more than 22 million discrete states

4) Results:

- Significantly reduced convergence time and increased rewards versus baseline DQN
- Successful deployment over a set of maps with varying density and complexity

5) Limitations:

- No temporal model employed (like LSTM)
- Static assumptions upon environments

D. RLPlanNav: DRL + Classical Planner

1) Problem Statement: Trajectory planning, is smooth and kinematically feasible, in unknown environments.

2) Methodology:

- Upper-level DRL (Recurrent Deterministic Policy Gradient) for sub-goal selection
- Lower-level classical planner (EGO-planner) for trajectory smoothing
- Use of RGB input from the simulated sensor suite

3) Innovations:

- Hierarchical control: DRL for decision-making, planner for control execution
- LSTM-enhanced DRL to escape local minima

4) Results: Outperforms both standalone planners and end-to-end DRL in terms of energy efficiency, path smoothness, and success rate.

5) Limitations:

- Tested on static obstacles only
- Lacks scalability to large-scale dynamic maps

E. GARTD: Adaptive Navigation with Attention and Velocity Constraints

1) Problem Statement: Adaptive altitude and speed control for UAVs in highly dynamic, cluttered, low-altitude 3D environments.

2) Methodology:

- Guide Attention Mechanism to switch focus between navigation and avoidance
- TD3 with LSTM
- Velocity-constrained actor loss to improve speed control

3) Innovations:

- Adaptive control in full 3D
- Custom reward and attention layers

4) Results: 14% increase in success rate, 14% reduction in collision rate over baselines.

5) Limitations:

- Computationally expensive
- Requires extensive tuning of attention mechanisms

F. Path Planning on Multi-Level Frontiers

1) Problem Statement: Efficient real-time path planning by uncharted front-based modelling under unknown environments.

2) Methodology:

- Multi-level Octree voxelization for frontier selection
- Historical State Record Tree (HSRT) for recovering from dead ends
- Adaptive voxel granularity according to environmental density

3) Innovations:

- Combines the advantages of frontier-based as well as sampling-based approaches
- Adaptive modelling saves up to 20% of memory and planning time.

4) Results:

- Great success rates in three complicated test environments.
- Planning is faster but less resource-consuming.

5) Limitations:

- Not learning-based, policy generalization is absent.
- Dynamic obstacles not considered.

IV. Comparative Analysis

The comparative evaluation has been planned to explain better UAV navigation and conflict resolution in an unknown environment. The analysis compares each of the methods along important dimensions such as learning paradigm, architecture of the network, perception and planning strategies, decision-making scheme, type of environment, and limitations observed.

A. Comparison of Methodology and Architecture

In Table III, approaches are summarized for each particular work. It can be identified from the table that hybrid methodologies, such as RLPlanNav, allow both learning- based reasoning and deterministic planning to achieve potentially smooth path generation, whereas other techniques mostly rely on learned policies. Enhancements using attention and memory (such as LSTM) as in GARTD3 indicate potential in dynamic environments.

B. The Environment: Level of Adaptability and Realism

These results suggest that GARTD3 and MARL-PPO excel in dynamic multi-agent environments while curriculums will result in better stability and convergence in structured yet complicated 3D airspaces. Classical planners remain computationally efficient, but lack adaptiveness based on learning.

TABLE III: Comparison of Methodologies and Architectures

Authors	**Learning Paradigm**	**Planner/Module**
Wang et al. (2024)	Deep Q-Learning	Rule-based Planner
Frémond et al. (2024)	Actor-Critic RL	Decentralized Coordination
Carvalho et al. (2023)	Value-based RL	Exploration-guided Planner
Xue and Chen (2024)	Hybrid DRL + Classical Planner	Low-level Deterministic Planner
Yin et al. (2024)	PPO with Temporal Features	Learned Policy Navigation
Duan et al. (2024)	Non-learning Heuristic	Greedy Exploration Planner
Carvalho et al. (2023)	Value-based DRL	NA

TABLE IV: Environment Complexity and Evaluation

Author	Environment	Key Challenges Addressed	Limitations
Wang et al. (2024)	Simulated 2D with visual obstacles	Visual perception, Obstacle Avoidance	No dynamic agents, high computation
Frémond et al. (2024)	3D Simulated Urban (Conflict Resolution)	Multi-agent collision avoidance, tactical decisions	Perfect communication assumed
Carvalho et al. (2023)	3D Grid-based City Model	Sparse rewards, increasing complexity	No temporal modeling
Xue and Chen (2024)	3D Forest Map with Obstacles	Safe path generation, Hierarchical decision-making	Lacks dynamic obstacle adaptation
Yin et al. (2024)	3D Simulated with Dynamic Obstacles	Real-time control, velocity safety, temporal awareness	High hyperparameter sensitivity
Duan et al. (2024)	Voxelized 3D Exploration Space	Efficient exploration, dead-end recovery	Static rule-based, lacks adaptability
Carvalho et al. (2023)	Complex 3D with task difficulty scheduling	Generalization in deep spaces, reward shaping	Not tested in dynamic multi-agent setups

TABLE V: Performance Evaluation of DRL-based UAV Navigation Methods

Author	Success Rate (%)	Collision Rate (%)	Avg. Episode Length
Wang et al. (2024)	92.5	5.3	210 steps
Fremond et al. (2024)	89.1	3.8	185 steps
Carvalho et al. (2023)	86.7	7.2	195 steps
Xue and Chen (2024)	90.4	4.9	170 steps
Yin et al. (2024)	93.2	3.5	200 steps
Duan et al. (2024)	88.5	6.0	230 steps

subsectionPerformance Benchmarking: A Synopsis To summarize the impacts of performance on these methodologies, Table V shows some of the key metrics across the six approaches. While RLPlanNav performed at the highest success rate in a static environment, GARTD3 and MARL-PPO exhibited good performance in dynamic and uncertain situations. The forward data deposit mechanism in FRDDM-DQN accelerates convergence but adds to complexity.

C. Key Observations

- Hybrid zones (e.g., RLPlanNav) mix flexibility from DRL with reliability from classical planning, yet fail to adapt in truly dynamic settings.
- Memory and attention-augmented DRL, GARTD3, excelled in robust real-time control under dynamic uncertainties.
- MARL-PPO systems for multi-agent utilization of airspace must be hardened against disruptions of communication.
- Curriculum learning systems exhibit superior learning curves but require a further infusion of temporal awareness to truly account for dynamic applications.
- Non-learning planners such as the Frontier method are very fast and easy to implement but generalize poorly underunknown and adversarial conditions.

Hence, the comparative analysis indicates that while the DRL-based systems offer the utmost adaptability in behaviour and autonomy, their real-world applicability for UAVs remains contingent upon improvements in the efficiency of computation, robustness against partial observability, and dynamic engagements in multi-agent environments. The combination of curriculum-learning, memory models, and hybrid planning offers the most promising path for upcoming UAV navigational systems.

V. Discussion and Future Directions

The integration of DRL into UAV navigation systems represents a significant step towards in achieving full autonomy. However, a few key challenges remain unsolved:

- Real-World Deployment: Many models are yet to be validated on real UAV hardware in dynamic environments.
- Memory & Computation: LSTM, attention mechanisms, and policy gradients require high computational resources for training and have high inference time.
- Safety Assurance: There is a lack of any formal safety verification framework for learned policies.

Future research directions include:

- Lifelong learning approaches would enable drones to continuously improve their navigation capabilities throughout their operational lifespan, rather than relying solely on pre-deployment training.
- Federated learning could revolutionize how drone fleets learn collectively while maintaining data privacy and reducing communication overhead.
- Incorporating formal safety guarantees into DRL navigation systems would address critical regulatory and practical concern.

VI. Conclusion

In-depth analysis of how Deep Reinforcement Learning and hybrid techniques are applied to drone navigation problems in unpredictable environments are reviewed across the literature. The paper encompasses a myriad of methodologies, extending from basic DQNs for single-drone scenarios to highly sophisticated multi-agent systems with implementations of PPO algorithms and memory-aware actor-critic architectures. Despite these works, there seems to be a grand divide that still separates simulation performance from real-world deployment. Persistent challenges are being encountered while transitioning from controlled test environments to real-world operations:

- The tracking of dynamic environments, where ever-changing conditions may disrupt the agents.
- Adhering to carefully defined safety constraints is a necessity.

- Computation on resource-constrained drone hardware is still limited.
- Ensuring consistent performance over various environmental conditions.

Theoretical aspects have great promise, but not able to solve many practical deployment issues that need thorough research that is still to be conducted before these systems can be deployed in mission-critical scenarios. Hybrid approaches combining classical navigation guarantees with DRL approachability are the most promising.

REFERENCES

[1] F. Wang et al., "Deep-reinforcement-learning-based UAV autonomous navigation and collision avoidance in unknown environments," Chinese Journal of Aeronautics, vol. 37, no. 3, pp. 237-257, 2024.

[2] R. Fremond et al., "Adaptive Multi-Agent Reinforcement Learning Solver for Tactical Conflict Resolution," IEEE Transactions on Aerospace and Electronic Systems, 2024.

[3] K. Carvalho et al., "UAV Navigation in 3D Urban Environments with Curriculum-based Deep Reinforcement Learning," ICUAS, 2023.

[4] Y. Xue and W. Chen, "Combining Motion Planner and Deep Reinforcement Learning for UAV Navigation in Unknown Environment," IEEE Robotics and Automation Letters, vol. 9, no. 1, 2024.

[5] Y. Yin et al., "Autonomous UAV Navigation with Adaptive Control Based on Deep Reinforcement Learning," Electronics, vol. 13, no. 2432, 2024.

[6] P. Duan et al., "Multi-Level-Frontier Empowered Adaptive Path Planning for UAVs in Unknown Environments," ICAIRC, 2024.