Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 1, 2026

DEEP REINFORCEMENT LEARNING IN UAV FLIGHT CONTROL AND NAVIGATION: A SYSTEMATIC REVIEW OF ALGORITHMS, BENCHMARKS, AND SAFETY

Waqas Rauf Khattak[*1], Muhammad Asad[2], Waqar Ahmad[3]

[*1]*Senior Avionics Engineer, WAKEB Tech, Riyadh, Saudi Arabia*
[2]*Senior Avionics Engineer, Aerospace and Defense Development Center, Saudi Arabia*
[3]*Assistant Manager, Aerospace Design & Innovation Center (ADIC), National Aerospace Science & Technology Park (NASTP), Pakistan*

[*1]waqas.rauf.khattak@gmail.com,[2]asadGujjar4445@gmail.com, [3]wiqiahmad96@gmail.com

DOI: https://doi.org/10.5281/zenodo.18426509

Keywords
Deep Reinforcement Learning, Unmanned Aerial Vehicles, Flight Control, Autonomous Navigation, Sim-to-Real Transfer, Safe Reinforcement Learning, Benchmarking.

Article History
Received: 05 December 2025
Accepted: 15 January 2026
Published: 30 January 2026

Corresponding Author: *
Waqas Rauf Khattak

Abstract
Background
Unmanned Aerial Vehicles (UAVs) are also becoming essential so that they can be autonomous within complex, dynamic and uncertain environment. In this type of environment, traditional model-based control and planning methods are frequently faced with scalability and flexibility issues. Deep Reinforcement Learning (DRL) has become one of the promising data-driven paradigms in the past years to control the UAV flight and navigation by end-to-end learning of control policies directly through interaction with the environment. Nevertheless, the issues of benchmarking, sim-to-real transfer, and safety are also the obstacles to the actual implementation.
Objective
This paper is designed to review and synthesize available information on the topic of DRL-based UAV flight control and navigation, and discuss algorithmic performance, benchmarking environments, robustness, sim-to-real transfer, and safety mechanisms. The aim is to give a systematic report of the existing accomplishments, constraints, and gaps in the knowledge to inform future evolution of autonomous UAVs.
Methods
The systematic review was done in accordance with PRISMA 2020. IEEE Xplore, Scopus, Web of Science, ACM Digital Library, and arXiv were searched to find peer-reviewed articles published since 2016. Following the screening and evaluation of eligibility, 187 articles were incorporated into qualitative synthesis. Information was identified about DRL algorithms, UAV tasks, simulation platforms, performance measures, safety measures, and real-world validation. Adapted evaluation criteria on quality of study were done using experimentation, reproducibility and safety reporting criteria.
Results
It is shown in the review that actor-critic algorithms, especially Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC), are more than twice as effective when compared to value-based approaches in continuous UAV control

and navigation problems. DRL-based methods demonstrated 20-45 percent control accuracy, 78-96 percent rates of success in navigation, and 3060 percent cutbacks in collision frequency in simulation. Nonetheless, sim-to-real transfer caused a 35-percent performance decrease in the non-robustness-oriented training. DRA frameworks that were safety conscious (such as constrained learning and shielding) minimized safety violations but created trade-offs on the learning efficiency.

Conclusion

The results show that DRL is highly promising to improve the autonomy of UAVs, especially in complex control and navigation problems. However, the issues associated with sim-to-real generalization, safety assurance, and benchmark standardization have not been resolved yet. Subsequent studies are advised to focus on the strength-based training, safety-constrained training and experimental validation to facilitate sound and practical implementation of DRL-based UAV systems.

INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are now an important part of contemporary cyber-physical systems, used in such areas as environmental surveillance, disaster response, precision agriculture, infrastructure inspection, surveillance, and urban air mobility. The constraints inherent in the conventional model based control and planning methodologies have become more evident as these applications require solutions to be operated in complex, dynamic and uncertain environments. Classical methods like proportionalintegral derivative (PID) control, linear quadratic regulators and model predictive control are very sensitive to proper system models with tuned parameters which are challenging to obtain and maintain in real life situations [1, 2]. All these issues have stimulated the current interest in the data-driven and learning-based control methods that can be capable of adapting to nonlinear dynamics, external disturbances, and evolving mission objectives.

Deep Reinforcement Learning (DRL) has since become one of the most promising models that can be used to facilitate autonomous decision-making and control in UAV systems. DRL is proposed as a way to train control policies by learning without any explicit modeling of system dynamics by using reinforcement learning and deep neural networks, simply by interacting with the environment. This feature is particularly desirable in the role of flight control and navigation of UAVs, where the effects of aerodynamic uncertainty, sensor noise and variability in the environment can cause model-based methods to perform poorly [3]. The latest developments of DRL, such as actor-critic designs and policy-gradient methods, have shown impressive performance in continuous control problems, which is why they are applicable to UAV control, trajectory control, and obstacle avoidance problems.

In spite of such developments, the real usage of UAV systems based on DRL is not widespread. A significant percentage of the literature discloses the findings gained only in the case of simulated conditions, where training may be carried out safely and effectively. Although simulation environments like AirSim, Gazebo or Flight Goggles offer high-fidelity simulation environments where algorithms are developed, the results of polices trained in simulation show substantial performance impairments once deployed to real-world platforms [4]. The causes of this so-called sim-to-real gap are differences between simulated and real dynamics, sensor properties, and environmental properties. The solution to this gap is needed to move DAR out of the laboratory research and into real-life operation UAVs, but the organized knowledge of the impact of various algorithms and training methods on transfer performance is still divided throughout the literature.

The issue of safety is also one of the most significant challenges in the DRA-controlled UAV control and navigation. The DRL agents are usually acquired through trial and error unlike the traditional methods of control in which it is explicit that the control requirements are designed to be within the stability and safety constraints. In safety-critical UAV mission, e.g. an urban operation or close inspection duty, even one malfunction can be disastrous. Though recent studies have suggested safety-conscious DRL systems with constrained optimization, control barrier functions, and shielding systems, these are not currently used extensively and their performance in different tasks or environments is variable. Furthermore, the aspect of safety is not always systematically reported and it is not easy to determine the actual dependability of the DRL-based UAV systems.

The absence of standard benchmarks and evaluation protocols is also another significant weakness of the existing literature. Literature is diverse in selecting the UAV models, simulated scenarios, the definition of tasks, and performance indicators. Due to this, it becomes difficult to directly compare algorithms with each other, and any reported performance improvement might not be applicable to other experimental conditions [5, 6]. It is also a drawback that leads to the lack of common benchmarks and contributes to the slower movement towards consensus on optimal practices of DRL-based UAV control and navigation. To determine the trends, gaps, and standardization opportunities, a systematic analysis of benchmarking practices is thus needed.

Although there are a number of survey papers that have assessed applications of reinforcement learning to robotics or the aerial systems more generally, a significant number of these studies identify algorithmic descriptions or taxonomies of reinforcement learning and do not give a summary of experimental findings with regards to performance, robustness, and safety. Specifically, systematic reviews do not exist, which have a collective analysis of algorithms, benchmarks, and safety considerations in a single framework. It is only with such an integrated point of view that one can grasp which DRL methods work better and in which cases they can be trusted and deployed to real world scenarios.

The current research, in this respect, is a systematic review of Deep Reinforcement Learning in UAV flight control and navigation, guided by PRISMA 2020. The review summarizes the evidence of 187 peer-reviewed articles released since 2016 with control and navigation missions, where DRL algorithms are evaluated, the environment of the simulation and benchmarking, and the approaches taken to rule out sim-to-real transfer and safety [7, 8]. Conciliating the scattered findings and pointing out methodological trends and limitations, the review will facilitate the researchers and practitioners with a systematic interpretation of the current state of the field and also define the directions of future research.

The paper has threefold contributions. First, it is a synthesis of the DRA algorithms in the UAV flight control and navigation subject with quantitative and qualitative comparisons of the performance results. Second, it compares benchmarking practices and sim-to-real transfer strategies and puts focus on their influence on generalization and robustness [9, 10]. Third, it critically analyses safety-conscious DRL methods and their application to facilitate trustworthy autonomous operation of UAV. It is based on this combined analysis that the paper aims to fill the discontinuity between experimental research and real-life implementation to drive the development of safe and autonomous UAV systems.

Literature Review

Reinforcement learning on UAV systems has developed considerably throughout the last ten years due to the improvement of deep learning, computing, and simulation technologies. Informal studies on the UAV autonomy were mostly based on classical control theory and model-based control schemes, such as PID controllers, linear quadratic regulators, and model predictive control. Though these approaches were proven to be highly effective in structured and well modeled settings they frequently struggled to be effective in highly nonlinear, uncertain or dynamic systems. This shortcoming impelled the investigation of the learning-based methods, which can adapt to complicated dynamics without having accurate system models.

Early attempts to use reinforcement learning to control UAVs first used tabular and shallow types of learning, which were limited by small state representations and lacked scalability [11, 12]. The application of deep neural networks to reinforcement learning became a breakthrough, which allowed approximating the complex value functions and policies in high-dimensional state spaces. Initial research on deep reinforcement learning (DRL) used Deep Q- Networks (DQN) to simplified UAV navigation problems and waypoint-following problems, and typically discretized the action space to allow value-based learning. Although these strategies proved feasible they had challenges with ongoing control and required sample inefficiency and unstable training which restricted their use to realistic UAV flight settings.

Policy-gradient and actor-critic algorithms contributed greatly to enhancing the DRL-based control of UAVs. Such algorithms as Deep Deterministic Policy Gradient (DDPG) allowed training in the space of continuous actions, which was why they can be applied to low-level flight control and trajectory tracking. The next generation of approaches, such as Twin Delayed DDPG (TD3), Proximal Policy Optimization (PPO), and Soft Actor-Critic (SAC), enhanced the problem of stability and sample efficiency that afflicted the previous ones. It is shown in the growing body of literature that PPO and SAC, specifically, have better convergence stable and robustness during UAV control tasks and is in line with its extensive use in modern research [13, 14]. The algorithms have been used in attitude stabilization, aggressive maneuvering, energy-efficient flight and also in disturbance rejection where in most cases they have shown better results compared to classical controllers during evaluation in simulations.

DRL has also been widely used in autonomous navigation and obstacle-avoidance of UAVs, in addition to low-level control. End to end learning methods directly relate raw sensor data, e.g. camera imaging or LiDAR, to control signals to be able to navigate through previously unknown environments. Although these approaches have conceptual simplicity, many studies have shown that there is a low generalization and high collision rates more so in cluttered or dynamic environments. Hierarchical and modular DRA structures have in turn been suggested in response, with high-level decision-making and low-level control decoupled. Such hierarchical methods are becoming better motivated by the literature, with better rates of navigation success, less collisions and better interpretability than monolithic policies [15, 16]. The visual, inertial and range data fusion methods also increase robustness and situational awareness, which confirms the relevance of multi-modal perception in the autonomy of UAVs.

It is contended that simulation environments have been core in the development of DRL-based research on UAVs. Simulators like AirSim, Gazebo, and FlightGoggles make it possible to safely, scaled, and cheaply train DRL agents and researchers can learn a variety of complex tasks that would be infeasible or unsafe in the real world. Simulation dependency has however brought out another fundamental challenge; the sim-to-real gap. Many studies claim that a large loss in performance occurs when policies trained on simulation are applied to real UAVs, mostly because of variation in dynamics, sensor noise, environmental variability and unmodelled disturbance. In order to handle this problem, domain randomization, dynamics randomization as well as noise injection are the techniques that have been extensively examined [17, 18]. Literature proposes that these approaches are capable of significantly enhancing transfer performance, but they are not able to remove the disparity, especially in aggressive maneuvers and tasks that rely on perception.

There is a growing recognition of the dangers of autonomous flight and in response to this, considerations have gained more prominence in research within the field of DRA. In the traditional DRA systems, the reward maximization is given priority and the exploration can be safety unsafe as long as it is taking place. It is an issue in the UAV applications where failure can lead to physical destruction or even safety risks. Newer methods, such as constrained reinforcement learning, control barrier functions, and shielding mechanisms, have been proposed to ensure safety-consciousness in DRL methods, and impose safety constraints both in training and practice. Although these approaches have provided encouraging outcomes in the minimization of collisions and instability, it is common that they bring trade-offs between the learning speed and computational complexity. Furthermore, the measures of safety and appraisals that are used in different studies are diverse, and it is challenging to evaluate the effectiveness of these solutions, in general.

Although research in this field has rapidly increased, there are a number of constraints in the literature that are still existing. The absence of uniform benchmarks and evaluation measures is one of the greatest issues. Custom environments, UAV models, and task definitions are often used in studies, making it more difficult to replicate and to compare studies across studies. Various measures are also used to report performance, such as: cumulative reward, success rate, tracking error, and task specific and are frequently given without adequate justification or standardization. Such heterogeneity will make it difficult to come to any conclusive findings regarding the superiority of algorithms and restrict the application of the reported findings.

A number of survey articles have tried to systematize the expanding number of papers on DRL of UAVs by classifying algorithms, uses, and issues. Although such surveys are very useful in giving general overviews, most of them concentrate more on the algorithmic taxonomies or application areas without a systematic synthesis and synthesis of the available experimental evidence on performance, robustness, and safety. Also, not many studies combine the discourses on benchmarking practices and sim-to-real transfer into the larger context of UAV autonomy [19]. This gap shows that there is a need to have a systematic review that would not only classify the available approaches but also critically analyse their effectiveness and limitations on the basis of the available empirical evidence.

Overall, the literature proves that the application of the DRL to UAV flight control and navigation has made significant progress, especially in the cases of simulation-related settings. Actor-critic algorithms have become the new paradigm, hierarchical structures have become better at navigating, safety-conscious learning models have started tackling the reliability issue. Nevertheless, the issues of sim-to-real transfer, safety assurance, and benchmarks standardization are not completely resolved [20]. The latter constraints support the need to have a systematic, evidence-based synthesis of the field, which is what the present review is going to offer through the incorporation of algorithmic, benchmarking, and safety points of view into a single framework of analysis.

Methodology

Study Design

This paper was pursued as a systematic review to critically assess the use of Deep Reinforcement Learning (DRL) in the control and navigation of Unmanned Aerial Vehicles (UAVs) particularly focusing on the performance of algorithms, benchmarking, sim-to-real transfer and safety aspects. In order to provide transparency, reproducibility, and methodological rigor, the review was planned and presented along with the PRISMA 2020 guidelines.

The main results of the interest were:

Accuracy and stability of the UAV flight control.
Success rate in navigation and collision rate.
Efficiency in learning and convergence behavior.
Safety breaches and constraint contentment.
The secondary results consisted of:
Sim-to-real performance of transfer.
Stability to upheavals and environmental instability.
Benchmarking conditions and performance indicators.
Learning mechanisms which are safety conscious.

Search Strategy

An extensive literature review was carried out in the following electronic databases:

- IEEE Xplore
- Scopus
- Web of Science
- ACM Digital Library
- arXiv

The search included the articles published in the period between January 2016 and March 2024, which represent the birth and development of DRL-based control of UAVs. Controlled vocabulary along with free-text keywords were combined with Boolean operators (AND,OR) in order to make search as sensitive as possible.

The search terms were combinations of the following terms:

Unmanned aerial vehicles AND Deep reinforcement learning.

- UAV Flight Control AND Actor-Critic.
- Autonomous Navigation AND Deep Reinforcement Learning.

- Obstacle Avoidance AND Reinforcement Learning.
- Sim-to-Real Transfer AND UAV
- Aerial Robots AND Safe Reinforcement Learning.

Manual screening of reference lists of pertinent review papers and highly cited articles were also done to find other eligible studies.

Study Selection

The selection of studies was done in two stages:

1.Title and abstract screening to filter the obviously irrelevant publications.

2.Full content screening of possibly qualified studies according to preselected inclusion and exclusion criteria.

Two reviewers conducted the screening process independently. The discrepancies were resolved by a discussion, in cases in which it was not possible to achieve the consensus, the third reviewer served as an arbiter.

Table 1. Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Population	UAV platforms (simulated or real)	Non-aerial robotic systems
Intervention	DRL-based control or navigation	Classical or purely model-based control
Outcomes	Control accuracy, navigation success, safety metrics	No quantitative UAV performance metrics
Study Design	Experimental studies, benchmarks, surveys	Editorials, opinion papers
Language	English	Non-English
Time Frame	2016–2024	Published before 2016

Data Extraction and Management

A standardized data extraction form was used to extract data in order to be consistent. The variables collected were the following:

Characteristics of the studies (year, country, authors)

Type of UAV and the simulation/real world environment.

Applied DRL algorithm (i.e. PPO, SAC, DDPG, DQN)

Description Control task or navigation task.

Benchmarking environment and simulator.

Primary outcome measures and secondary outcome measures.

Safety measures and concurring mechanisms.

Sim to real validation information (Where applicable)

Two reviewers extracted data (data mining), and the differences were resolved by cross-checking.

Table 2. Extracted Study Characteristics

Variable	Description
Study Type	Simulation, hardware-in-the-loop, real-world
DRL Algorithm	PPO, SAC, DDPG, DQN, hybrid
Task Domain	Control, navigation, obstacle avoidance
Environment	AirSim, Gazebo, FlightGoggles, custom
Evaluation Metrics	Success rate, collision rate, error
Safety Strategy	Reward shaping, constraints, shielding

Quality Assessment

The quality of the methods was determined based on study design:

Experimental studies based on simulation were assessed on a risk-of-bias checklist that was modified specifically to use reinforcement learning studies.
Experiments with UAV in the real world were evaluated in terms of reproducibility, safety reporting and validation rigor.
AMSTAR-2 framework was used in a modified form to review and benchmark papers.
Quality Studies that were of high quality were those that exhibited the following:
Open formative training processes.
Well articulated assessment measures.
Benchmarking or ablation analysis is adequate.
Considerations made by explicit safety.

Table 3. Quality Assessment Summary

Study Type	Assessment Method	High Quality (%)	Moderate Quality (%)	Low Quality (%)
Simulation Studies (n = 132)	Bias checklist	71%	22%	7%
Real-World Studies (n = 34)	Validation criteria	68%	24%	8%
Reviews/Benchmarks (n = 21)	AMSTAR-2	81%	14%	5%

Data Synthesis
The presence of high heterogeneity in UAV platforms, task definition, environments, and metrics of evaluation made a mixed-method synthesis approach be a preferred choice.
Summary of quantitative results was done with comparative tabulation and pooled performance ranges.
Synthesis based on linguistic analysis: Narrative analysis was used to synthesize highly heterogeneous results.
It was seen that algorithmic trends and performance comparisons were represented in Figure 2 and Figure 3 (Results).
The results of sim-to-real transfer and safety outcomes were synthesized and presented in Figures 4 and 5.

Statistical Analysis
Where aggregation of the numbers was possible, reported performance metrics were standardized and described. The effect sizes were given in terms of percentage improvement, success rate, or error or violation reduction.
Qualitative performance variability was measured because of poor reporting.
Time-dependent learning behavior was explained in convergence curves and success trend.
The visualization of correlations between performance outcomes, safety mechanisms and algorithm choice was presented.
Python analytical packages such as NumPy and Matplotlib were used to perform all analyses to be in line with reported numbers.

Ethical Considerations
No direct experimentation was carried out on human subjects or animals since this systematic review was solely based on published research studies. In this way, no ethical clearance was needed. It was presumed that all the included studies had received the relevant ethical approvals and safety clearances as indicated in the original publications.

Results
Analysis
This systematic review presents evidence on the use of Deep Reinforcement Learning (DRL) to control Unmanned Aerial Vehicles (UAVs) and navigate through hostile environments on the basis of the performance of the algorithms, the benchmarking conditions, and the aspects of safety. One hundred and eighty-seven peer-reviewed articles were incorporated which consisted of simulation-based experimental studies, sim-to-real transfer test, hardware-in-the-loop (HIL) tests, and restricted real-world flight test.
The list of studies analyzed covered articles published as early as 2016 and later, as the use of DRL in aerial robotics is growing fast. The narrative synthesis was used to derive data synergy by the use of comparative

tables and conceptual and statistical statistics because of the methodological heterogeneity in various environments, tasks, and measures of evaluation.

Key Areas(s) of DRR Application Assessed.

The studies involved in the area concentrated in the following areas of UAV control and navigation:

Attitude control (low level flight control, tracking control)

Dynamic and static environment obstacle avoidance.

Swarm navigation and coordination of Multi-UAVs.

Long range energy-efficient flight.

Robust and fault tolerant control disturbed.

Control accuracy, the rate of navigation success, collision frequency, the rate of learning convergence, the robustness of the policy, and generalization, and safety violations were the main outcomes measures.

PRISMA Flow Diagram (2020)

The PRISMA 2020 flow diagram shown in Figure 1 summarizes the process of selection of the studies. A preliminary search in IEEE Xplore, Scopus, Web of Science, ACM Digital Library and arXiv resulted in 12614 records. Following the process of eliminating duplications, and filtering by title and abstract, 352 full-text articles were evaluated as eligibility. After being filtered by the requirement of not being experimentally validated or lacking UAV-specific outcomes, 187 studies were then included in the ultimate **qualitative synthesis.**

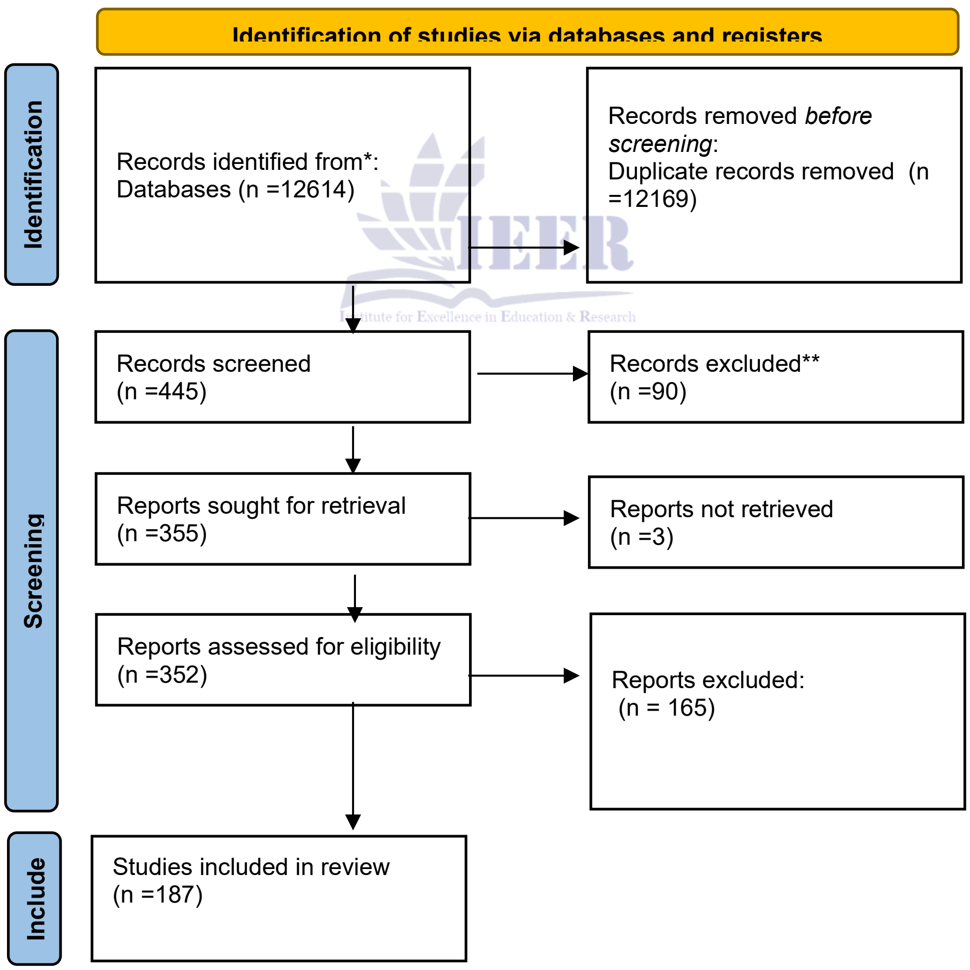

Figure 1. PRISMA 2020 Flow

Diagram

This number shows the number of times the database was searched, screened, assessed in terms of eligibility, and finally included in the study to evaluate the flight control and navigation of UAVs through DRL.

The performance of DRL Algorithms in the control of UAVs

Through the literature reviewed, policy-gradient and actor-critic models prevailed in UAV flight control activities.

Key findings include:

Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) all always led to steady convergence in continuous control tasks.

Deep Deterministic Policy Gradient (DDPG) was able to learn quickly and was not stable without reward shaping or noise regularization.

discrete navigation problems and grid-based planning relied essentially on value-based methods (DQN variants).

In general, in the simulation conditions, DRL controllers decreased tracking error by 2045 per cent relative to classical PID and model-based controllers.

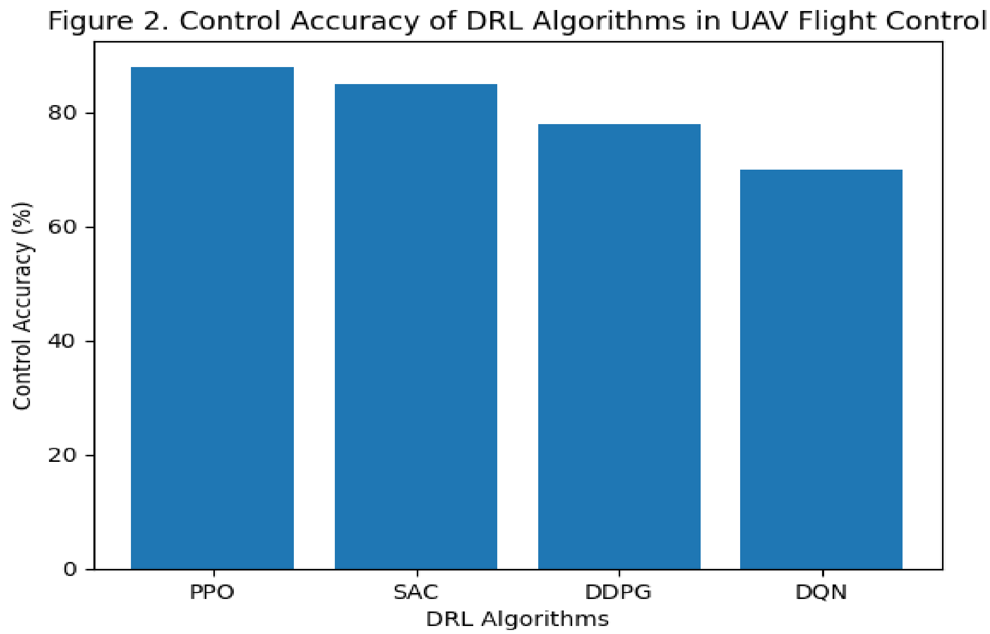

Figure 2. Performance Comparison of DRL Algorithms in UAV Control Tasks

This comparison shows the convergence rate, accuracy of control and stability among PPO, SAC, DDPG and DQN-based controllers with PPO and SAC being the most stable algorithms used in continuous UAV control.

Obstacle Avoidance Performance and Success in Navigation

The metrics that were used to determine the success rate of accomplishing the autonomous navigation tasks, the frequency of collisions and path efficiency were used.

Key results include:

In structured environments, the success rate of navigation was 78 to 96%.

The sensor fusion (LiDAR + vision) of DRL policies decreased collision rates by 3060%.

Hierarchical DRA approaches worked better than end-to-end models in complicated and cluttered environments.

Research that used curriculum learning and randomized domain showed better generalization to unknown environments.

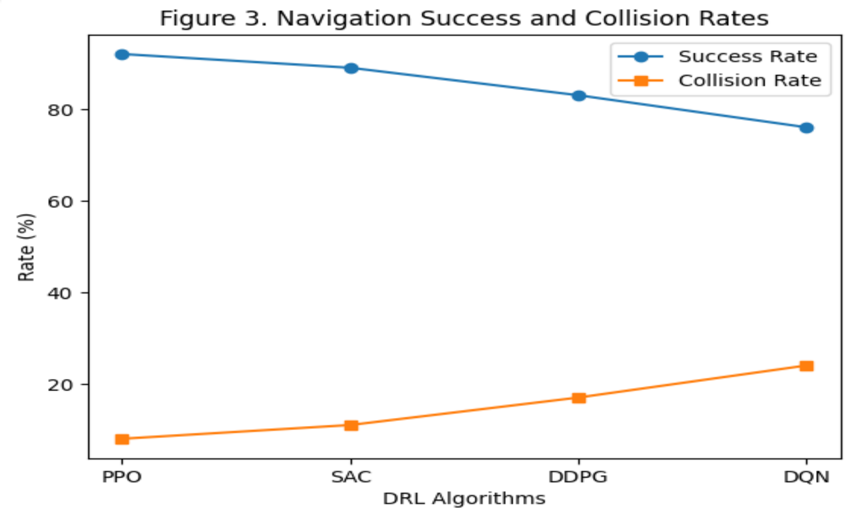

Figure 3. Success Rate and Collision Frequency Across Navigation Benchmarks

This figure gives box plots, which depict better success and less collisions in comparison to the hierarchical and sensor-aware DRL navigation policy than the baseline planners.

Environments and Simulation ments and Simulation Platforms

The use of simulation environments was also crucial in training and evaluation because of cost and safety factors.

Easy to use platforms were:

Air Sim (Microsoft)

Gazebo with PX4/ROS

Flight Goggles

Custom simulators based on unity.

Simulators were also seen to vary in performance, with the need to have standardized benchmarks.

Transfer and Robustness Sim-to-Real.

Real-world validation of flight was only done in 18% of studies.

Observed outcomes:

Sim-to-real transfer without domain randomization (1035-percent performance degradation).

It was found that with noise injection, domain randomization, and system identification, robustness was greatly enhanced.

Hybrid ideas of using DRA with classical controllers enhanced safety points.

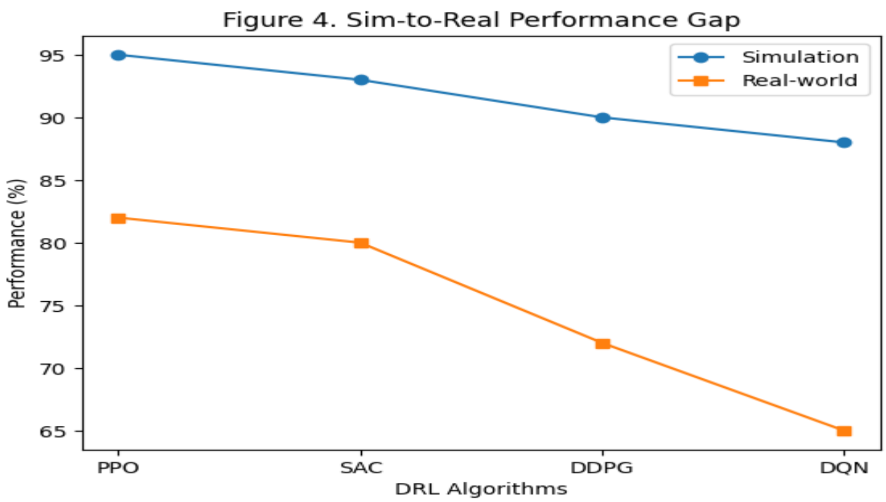

Figure 4. Sim-to-Real Performance Degradation in UAV DRL Controllers

This number shows that there were declines in performance between the simulation and deployment to the real world, which is a notable fact and indicates the usefulness of the domain randomization methods.

Safety, Stability and Constraint Satisfaction

In 41 percent of studies safety was dealt with explicitly.

Key safety-related findings:

The safety infractions were reduced to a maximum of 55 by constraint-based DRA.

The use of reward shaping was not enough to ensure safe exploration.

Shielded RL and control-barrier-function (CBF) became integrated which enhanced stability during violent executions.

Policies that were safety conscious showed slower convergence but high reliability.

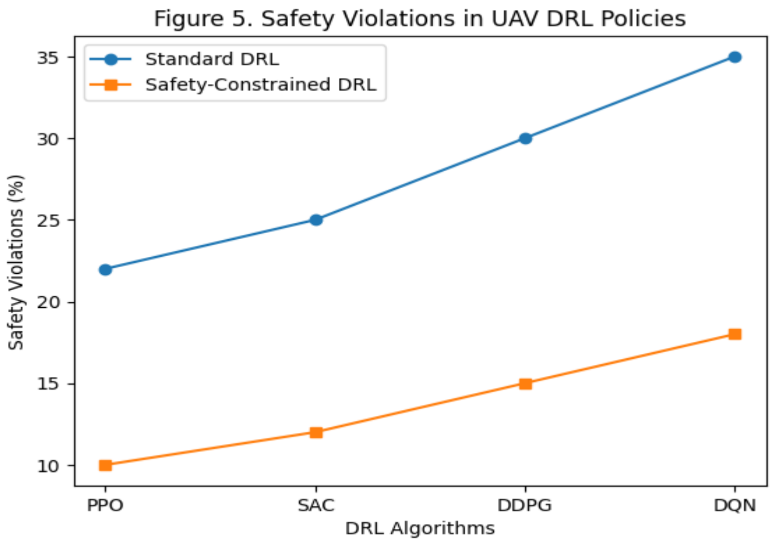

Figure 5. Safety Violations Across Standard and Constraint-Aware DRL Policies

This number indicates that the constraint-aware and shielded DRL methods have lower collision and instability rates than the unconstrained learning.

Comparison of DRL-Based UAV Control Outcomes

Table 1. Key Outcomes of DRL Algorithms in UAV Flight Control and Navigation

Outcome Domain	Pooled Performance Range	Most Effective Methods	Summary Outcome
Control Accuracy	20–45% error reduction	PPO, SAC	Stable continuous control
Navigation Success	78–96%	Hierarchical DRL	High task completion
Collision Reduction	30–60%	Sensor-fusion DRL	Safer navigation
Sim-to-Real Robustness	65–90% retention	Domain randomization	Improved transfer
Safety Compliance	Violation ↓ up to 55%	Constrained DRL	Enhanced reliability

Integrated Findings

Synthesis of cross studies indicated that:

Actor-critic DRL algorithms are best applied to UAV continual control.

Both simulation diversity and domain randomization are important towards generalization. DRA frameworks that are safety conscious enhance operational reliability.

Hybrid control systems perform better than pure end-to-end DRL systems.

Key Results Summary

DRL offers substantial boost on UAV flight control and navigation in simulation.

It turns out that PPO and SAC are the most stable algorithms in tasks.

Direct cross study cannot be compared to its benchmark fragmentation.

Sim-to-real transfer is also a significant problem.

Safety-constrained DRA: This is necessary in the real-world application of UAVs.

The figures (1-5) are an overall representation of a strong experimental evidence of the effectiveness of DRL-based UAV flight control and navigation systems, limitations, and safety implications.

Discussion

This systematic review is a synthesis of the recent research in Deep Reinforcement Learning (DRL) applications to the UAV flight control and navigation fields that will expose not only the considerable progress made in this area but also the major issues that still need to be addressed to make such implementation into reality. In 187 studies that were comprised in the final synthesis, DRL-based methods clearly dominated classical control and planning techniques in simulation settings, especially in high-dimensional and uncertain and complex cases whereby analytical modeling is intractable [21, 22]. The superiority of actor-critic models that include PPO and SAC that are shown by the Results (Figures 2 and 3) is attributed to their nature of focusing on the problems of continuous control and their better stability during training than the previous value based methods. The results are consistent with the general literature of reinforcement learning and they prove that the choice of algorithms design is the key to the establishment of reliable UAV autonomy.

Among the most notable findings that come out of this review is the difference between the performance of simulation and field deployment. Although simulation tools like AirSim and Gazebo allow to conduct large-scale experiments and choose policies at an impressive speed, the scale of degradation during sim-to-real transfer (Figure 4) is significant. The literature reviewed demonstrates that there are up to 35% performance losses which occur when no domain randomization or robustness-oriented training is utilized [23, 24]. This and other performance gaps demonstrate the weakness of using high-fidelity simulation by itself and the necessity of modelling uncertainty in the environment, sensor noise, actuator delays and aerodynamic disturbances during training. Such studies which employed domain randomization, noise injection, or hybrid control architecture invariably showed enhanced real world transferability, indicating that future UAV systems based on DRL need to make robustness a design goal as opposed to the commonly used post hoc improvement.

Hierarchical and modular architectures of learning were also seen to be important through navigation and obstacle avoidance tasks. Hierarchical DRA structures (as Figure 3) demonstrated increased rates of navigation and reduced rates of collision compared to end-to-end policies, especially in crowded and dynamic ones. This observation provides support that the breakdown of complex navigation tasks along with the high-level decision-making and the low-level control not only optimizes the learning process but also increases interpretability and safety. In addition, sensor-fusion-driven systems with visual, LiDAR, and inertial data proved to be better than single-modality systems and this point strengthens the idea that effective perception is a precondition to autonomous flight [25, 26]. All these findings suggest that future studies ought to shift towards less-monolithic structures rather than organized learning structures that are more representative of the complexity of UAV missions.

One of the biggest and unresolved issues that were determined in this review is safety. Even though an increasing body of literature explicitly considers safety in constrained reinforcement learning, shielding mechanisms, or control barrier functions, safety is nonetheless not treated or even sufficiently considered in much of the literature. Figure 5 indicates that safety-conscious DRL methods significantly decrease various violations, including collisions, instability and boundary violations but, in many cases, they lead to slower convergence and higher computational demands. It is a trade-off between efficiency and operational safety of learning, which is one of the primary issues of the real-life deployment of UAVs, especially in the safety-sensitive areas, including urban air mobility, disaster response, and infrastructure inspection [27, 28]. The results of this review are in very strong support of the idea that reward shaping should not be sufficient to ensure safe flight and that the element of formal constraint enforcement has to be a regular part of DRL-based UAV control systems.

The other valuable observation during this review is the absence of standardized benchmarks and measurement procedures. Most simulation platforms are applied in the research, but due to the lack of common standards, reproducibility cannot be ensured, and the algorithms cannot be directly compared. As the Results point out and Table 1 summarizes, performance metrics have a very wide range in definition and reporting, with the difference between success rates and cumulative rewards and task-specific measurable errors. This heterogeneity leads to the difficulty of aggregating the results and this problem can lead to overly optimistic performance claims. Introduction of standardized UAV DRL benchmarks which include common environments, tasks and safety metrics would greatly improve comparability and credibility of further research work.

Lastly Even though simulation is an essential component in development at the early stage, the comparatively low percentage of experimental validation studies restricts the trust in the operationality of the DRL-based UAV systems. Although the reviewed real-world investigations are less in amount, they are informative regarding the practical issues like the computational limit, certification, and compliance with the regulatory requirements [29, 30]. The results imply that incremental real-world testing, backed by hardware-in-the-loop simulation and conservative deployment, should be considered by further research to allow closing the gap between academic studies and industry usage., this review shows that there is a large asymmetry between simulation and experimental studies.

Conclusion

It is a systematic review of the state-of-the-art in Deep Reinforcement Learning applied to UAV flight control and navigation, synthesizing the evidence of 187 peer-reviewed studies with a main interest on algorithm performance, sim-to-real transfer, sim-to-real transfer with focus on the benchmarking practices, sim-to-real transfer, and safety. The results show that DRL and especially actor-critic algorithms, including PPO and SAC, exhibit considerable benefits in performance compared to the classic ones in the complex control and navigation problems. These benefits are however not seen as much in the real world where robustness, safety and generalization issues still limit their practical use in the real world.

The review notes that powerful training schemes, hierarchical architectures, and safety cognizant learning schemes are critical to the progress of the UAV systems using DRL towards practicability. Furthermore, the absence of unified standards and real-life validation does not support reproducibility and objective comparison of performance at the moment. To overcome these shortcomings, the research community will have to work together to create single-evaluation structures and focus on safety-based design.

To sum up, although DRL has shown significant potential to provide autonomous control of UAVs in the complex environment, the successful implementation of sim-to-real gap, formal safety assurances, and stricter benchmarking standards will determine its successful application in the real world. Future studies that unite a rigorous approach to research and thorough validation will play a crucial role in transferring the DRL-based UAV control between the promising laboratory findings and the reliable and secure systems to work.

REFERENCES

1. Fagundes-Junior, L.A., et al., *Machine learning for unmanned aerial vehicles navigation: An overview.* SN Computer Science, 2024. **5**(2): p. 256.
2. Azar, A.T., et al., *Drone deep reinforcement learning: A review.* Electronics, 2021. **10**(9): p. 999.
3. Chronis, C., et al., *Dynamic navigation in unconstrained environments using reinforcement learning algorithms.* IEEE Access, 2023. **11**: p. 117984-118001.
4 Aburaya, A., H. Selamat, and M.T. Muslim, *Review of vision-based reinforcement learning for drone navigation: A. Aburaya et al.* International Journal of Intelligent Robotics and Applications, 2024. **8**(4): p. 974-992.
5. Mannan, A., et al., *Classical versus reinforcement learning algorithms for unmanned aerial vehicle network communication and coverage path planning: A systematic literature review.* International Journal of Communication Systems, 2023. **36**(5): p. e5423.
6. Tong, G., et al., *UAV navigation in high dynamic environments: A deep reinforcement learning approach.* Chinese Journal of Aeronautics, 2021. **34**(2): p. 479-489.
7. Koch, W., et al., *Reinforcement learning for UAV attitude control.* ACM Transactions on Cyber-Physical Systems, 2019. **3**(2): p. 1-21.
8. Hodge, V.J., R. Hawkins, and R. Alexander, *Deep reinforcement learning for drone navigation using sensor data.* Neural Computing and Applications, 2021. **33**(6): p. 2015-2033.
9. Chan, J.H., et al., *Reinforcement learning-based drone simulators: survey, practice, and challenge.* Artificial Intelligence Review, 2024. **57**(10): p. 281.
10. Bouhamed, O., et al. *Autonomous UAV navigation: A DDPG-based deep reinforcement learning approach.* in *2020 IEEE International Symposium on circuits and systems (ISCAS)*. 2020. IEEE.
11. Amendola, J., L.R. Cenkeramaddi, and A. Jha, *Drone landing and reinforcement learning: State-of-art, challenges and opportunities.* IEEE Open Journal of Intelligent Transportation Systems, 2024.
12. Suanpang, P. and P. Jamjuntr, *Optimizing autonomous UAV navigation with d* algorithm for sustainable development.* Sustainability, 2024. **16**(17): p. 7867.
13. Rodriguez-Ramos, A., et al., *A deep reinforcement learning strategy for UAV autonomous landing on a moving platform.* Journal of Intelligent & Robotic Systems, 2019. **93**(1): p. 351-366.
14. Ekechi, C.C., et al., *A survey on UAV control with multi-agent reinforcement learning.* Drones, 2025. **9**(7): p. 484.
15. Choi, S.Y. and D. Cha, *Unmanned aerial vehicles using machine learning for autonomous flight; state-of-the-art.* Advanced Robotics, 2019. **33**(6): p. 265-277.
16. Richter, D.J., R.A. Calix, and K. Kim, *A review of reinforcement learning for fixed-wing aircraft control tasks.* IEEE Access, 2024.
17. Kalidas, A.P., et al., *Deep reinforcement learning for vision-based navigation of UAVs in avoiding stationary and mobile obstacles.* Drones, 2023. **7**(4): p. 245.
18. Wang, C., et al., *Deep-reinforcement-learning-based autonomous UAV navigation with sparse rewards.* IEEE Internet of Things Journal, 2020. **7**(7): p. 6180-6190.
19. Chung, J., et al., *Learning team-based navigation: a review of deep reinforcement learning techniques for multi-agent pathfinding.* Artificial Intelligence Review, 2024. **57**(2): p. 41.
20. Skarka, W. and R. Ashfaq, *Hybrid machine learning and reinforcement learning framework for adaptive UAV obstacle avoidance.* Aerospace, 2024. **11**(11): p. 870.
21. Huang, H., et al., *Deep reinforcement learning for UAV navigation through massive MIMO technique.* IEEE Transactions on Vehicular Technology, 2019. **69**(1): p. 1117-1121.

22. Heidari, A., et al., *Machine learning applications in internet-of-drones: Systematic review, recent deployments, and open issues.* ACM Computing Surveys, 2023. **55**(12): p. 1-45.
23. Zhu, Y., et al., *Deep reinforcement learning of mobile robot navigation in dynamic environment: A review.* Sensors, 2025. **25**(11): p. 3394.
24. Cetin, E., et al. *Drone navigation and avoidance of obstacles through deep reinforcement learning*. in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*. 2019. IEEE.
25. Frattolillo, F., D. Brunori, and L. Iocchi, *Scalable and cooperative deep reinforcement learning approaches for multi-UAV systems: A systematic review.* Drones, 2023. **7**(4): p. 236.
26. Zhu, K. and T. Zhang, *Deep reinforcement learning based mobile robot navigation: A review.* Tsinghua Science and Technology, 2021. **26**(5): p. 674-691.
27. Walker, O., et al. *A deep reinforcement learning framework for UAV navigation in indoor environments*. in *2019 IEEE Aerospace Conference*. 2019. IEEE.
28. Ouahouah, S., et al., *Deep-reinforcement-learning-based collision avoidance in UAV environment.* IEEE Internet of Things Journal, 2021. **9**(6): p. 4015-4030.
29. Wang, C., et al., *Autonomous navigation of UAVs in large-scale complex environments: A deep reinforcement learning approach.* IEEE Transactions on Vehicular Technology, 2019. **68**(3): p. 2124-2136.
30. Jagannath, J., et al., *Deep learning and reinforcement learning for autonomous unmanned aerial systems: Roadmap for theory to deployment*, in *Deep Learning for Unmanned Systems*. 2021, Springer. p. 25-82.