vehicles

MDPI

Article

Simulation-Driven Approach to Evaluate a Reinforcement Learning-Based Navigation System for Last-Mile Drone Logistics

Zakaria Benali and Amina Hamoud *

School of Engineering, University of the West of England, Bristol BS16 1QY, UK; zakaria2.benali@live.uwe.ac.uk
* Correspondence: amina.hamoud@uwe.ac.uk

Abstract

Unmanned Aerial Systems (UAS) offer sustainable solutions for urban last-mile logistics, yet existing navigation algorithms struggle with the complexity of dynamic metropolitan environments. This study optimises a reinforcement learning (RL)-based guidance, navigation, and control (GNC) algorithm using a Proximal Policy Optimisation (PPO) model within a high-fidelity simulation of Bristol City Centre. The primary contribution is training the RL model to autonomously detect and avoid dynamic obstacles, specifically manned aircraft, to ensure safe and legal drone operations. Additionally, flight operations are continuously monitored via a Structured Query Language (SQL) database to verify compliance with low airspace regulations. Simulation results demonstrate that the proposed framework achieves high obstacle detection accuracy under nominal conditions, while the implementation of curriculum learning significantly enhances the system's adaptability and recovery capabilities during high-speed, dynamic encounters.

Keywords: proximal policy optimization; urban air mobility; collision avoidance; reinforcement learning; low airspace regulation; simulation environment

1. Introduction

Drone operations can significantly lower carbon emissions, ease traffic, and help achieve global sustainability goals. It is still quite difficult to guarantee their safe and independent operation in complex urban settings. The incapacity of autonomous drones to consistently navigate dynamic urban environments is a significant issue, particularly in the presence of moving obstacles like a manned aircraft. Recent studies shown in Figure 1 demonstrate that over a three-year period, 218 near-midair drone-manned aircraft collisions were documented, with 58 occurring in 2020, 82 in 2021, and 81 in 2022 [1].

When aviation authorities formally investigate these proximity events, they are assigned a risk classification evaluating collision potential. While some incidents pose a direct danger, many fall into other categories [2,3], such as:

- **Category B**: Safety Not Assured in a proximity event where the safety of the aircraft was potentially compromised [2,3].
- **Category C**: No Risk of Collision in a proximity event where a collision was never imminent, or any potential danger was successfully avoided [2].
- **Category D**: Risk Not Determined in a situation where a lack of data, or conflicting evidence, makes it impossible to assess the actual level of danger [2].

Check for updates

Academic Editor: Tuan Nguyen Gia

Received: 17 February 2026
Revised: 18 March 2026
Accepted: 28 March 2026
Published: 8 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

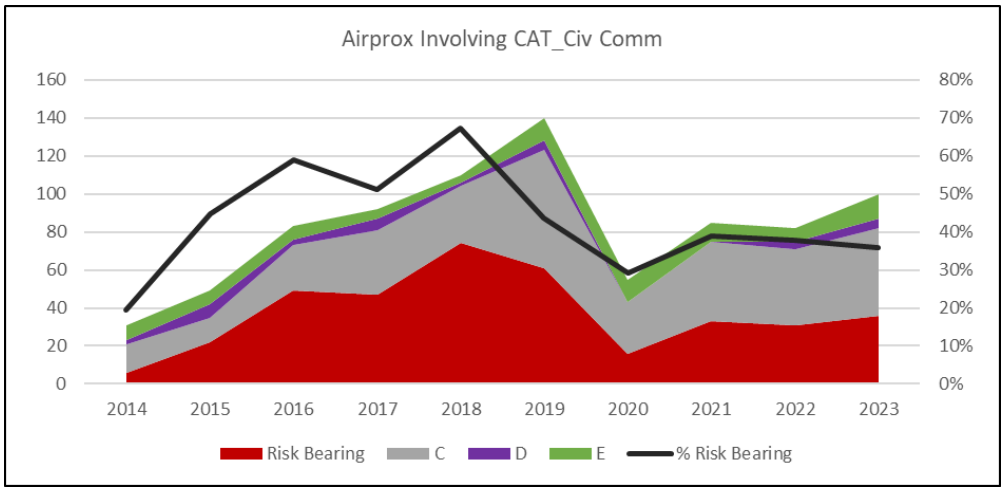


Figure 1. Commercial manned aircraft and drone encounters over the last decade [1]. (Note: This chart is reproduced directly from the UK Airprox Board. The left y-axis indicates the total number of incidents, and the right y-axis represents the percentage of risk-bearing incidents.).

Even accounting for encounters with no ultimate collision risk, the volume of incidents falling into categories where safety is unassured or indeterminable demonstrates a pressing issue.

By refining a reinforcement learning (RL)-based GNC system that can recognise and steer clear of dynamic obstacles in an urban simulation, this study seeks to solve this problem. The simulation employs artificial training data and this work's main contribution is the training and optimisation of a Proximal Policy Optimisation (PPO) RL model that has been taught to identify and steer clear of helicopters, enabling safer autonomous drone navigation. The system incorporates an SQL database that records drone flight data in connection to the Civil Aviation Authority's (CAA) altitude limitations (120 m) to guarantee partial compliance with regulatory frameworks. This promotes more consistent drone operation and makes it easier to work towards scalable legitimate drone deployment in the future. While these statistics underscore the urgent operational risks, a fundamental limitation persists in current Guidance, Navigation, and Control (GNC) systems. Traditional deterministic or rule-based GNC architectures are heavily constrained by their inability to adapt in real-time to unpredictable, high-speed dynamic obstacles such as emergency medical or police helicopters operating within complex urban geometries. Furthermore, acquiring the vast amounts of edge-case data required to train more robust, reactive avoidance systems is inherently unsafe in real-world settings. To directly address these limitations, this study proposes a simulation-driven approach to train and evaluate a reinforcement learning (RL)-based GNC system. By refining a Proximal Policy Optimisation (PPO) model within a high-fidelity urban simulation, this work enables the UAV agent to safely learn and execute adaptive, real-time collision avoidance strategies against dynamic manned aircraft.

2. Literature Review

It is essential to comprehend the status of drone navigation technology in an urban simulation situation. The following literature review highlights the use of synthetic data and urban simulation environments to overcome data shortages and safety constraints in autonomous drone training, while also examining the present complexity of various simulation environments and, more significantly, GNC algorithms.

2.1. Urban Simulation Environments for Drone Navigation

Training drones to autonomously perform complex tasks in dynamic environments demands extensive annotated datasets to develop robust navigation models [4]. Since data

availability is limited and gathering real-world data is costly, employing synthetic data from simulation environments to train GNC models for autonomous drones remains the preferred strategy [5]. Realistic and effective Unmanned Aerial Vehicle (UAV) testing environments are increasingly important as the demand for UAVs grows [6]. The "simulation-reality gap" has been introduced as a performance metric to evaluate and enhance model transfer from simulation settings to real-world conditions [4].

In this context, AirSim has been widely adopted to generate realistic datasets, focusing on bridging the gap between simulated and real-world scenarios using Simulation-to-Reality (Sim2Real) learning [7,8]. These studies emphasized the necessity of improving the realism of simulation environments and proposed feedback mechanisms to adapt to environmental factors observed in real-world tests. However, both are limited in modeling diverse and complex urban elements such as police and ambulance helicopters, which are essential for increasing the resilience and realism of drone modeling methods.

Similarly, other research has highlighted the importance of realistic simulation environments for efficient GNC algorithm development and testing [9,10]. For instance, the authors of these studies utilized Unreal Engine's AirSim to create distinct synthetic scenarios: Blocks and Landscape as testbeds for aerial navigation. This approach parallels subsequent work using AirSim within Unreal Engine to simulate urban neighborhood environments [10,11]. The integration of Gazebo with the Robot Operating System (ROS) to create simulation environments also reflects parallel advancements, particularly in regulatory applications such as modeling FAA Right-of-Way rules for drone interactions [11,12].

As a result, enhancing the environmental complexity and realism of UAV simulation environments has become a significant research priority. However, most existing studies remain limited in addressing dynamic obstacles and changing weather conditions [4,7,9–11]. Addressing these gaps requires continued progress in Guidance, Navigation, and Control (GNC) models, where advancements in autonomous decision-making, path planning, and flight control systems are critical for reliable drone maneuvering in complex urban contexts.

2.2. Reinforcement Learning Approaches to Guidance, Navigation & Control

One promising approach within GNC models is Reinforcement Learning (RL) as shown in Figure 2 , a machine learning technique in which an agent interacts with its surroundings to learn optimal decision-making strategies that maximize cumulative rewards [13]. RL has increasingly been applied in recent advancements in control and navigation algorithms for drone navigation in dynamic environments [14].

Several RL-based techniques have been proposed, each addressing specific aspects of UAV control and navigation, though they exhibit various limitations, as summarized in Table 1. For instance, a deep reinforcement learning (DRL) approach employing dynamic reward structures to balance exploratory and conservative behaviors achieved significantly improved path efficiency and obstacle avoidance in high-density, dynamic settings [15]. This method also embedded positional and angular changes into the state space to enhance decision-making. However, the use of sparse reward structures limited effective learning by providing feedback only when certain goals were achieved, which slowed convergence and hindered UAV training for dynamic urban navigation. Similarly, a hierarchical RL framework integrated visual odometry (VO)-based safe navigation by combining a high-level RL policy for waypoint generation with a low-level classical controller for command execution [16,17]. By utilizing semantic images, this approach enhanced safety in visually ambiguous zones, though challenges persisted in generalization to real-world conditions and in handling the high computational demands of training.

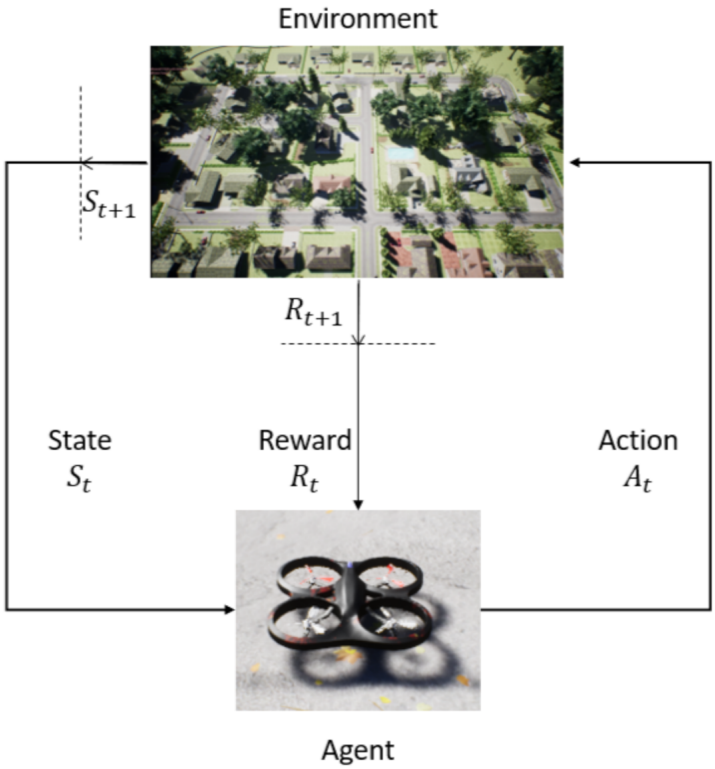


Figure 2. Reinforcement Learning Setup in a Simulation [10].

Table 1. Summary of RL Algorithms for UAV Navigation in Urban Environments.

Algorithm	Pros	Cons
Deep Reinforcement Learning (DRL) with Dynamic Rewards [15]	• Enhanced path planning in dense environments. • Good trade-off between exploration and exploitation while training.	• Slow convergence for sparse reward design. • Computationally intensive for complex environments.
Proximal Policy Optimization (PPO) with Curriculum Learning [16]	• Flexible and adaptive to hazardous environments. • Long Short-Term Memory (LSTM) integration improves navigation memory. • Efficient sensor data fusion enhances obstacle avoidance.	• High computational demands limit scalability. • Requires extensive fine-tuning of state representations. • Challenges in interpreting model decisions.
Hierarchical RL with Visual Odometry (VO-Safe RL) [17]	• Combines RL flexibility with traditional control robustness. • Semantic images reduce sim-to-real gap. • Enhances VO-based localisation performance.	• Computationally expensive to train. • Requires high-quality visual data for localisation. • Low accuracy in larger environments.
Target Following Deep Q-Network (TF-DQN) [18]	• Curriculum training adapts well to complex environments. • Robust obstacle avoidance and tracking accuracy. • Simplifies navigation with discretised control.	• Relies heavily on high-quality sensor data. • Limited in handling real-time decision-making. • Performance degrades with noisy or dynamic sensor inputs.

Table 1 summarises the advantages and limitations of several promising RL-based techniques for urban UAV navigation. Among these, the Target Following Deep Q-Network (TF-DQN) introduced a curriculum training strategy to gradually adapt UAVs to complex scenarios, focusing on persistent target tracking in urban environments [18]. The use of discrete control commands simplified navigation while maintaining strong obstacle avoidance, though the method struggled with real-time decision-making in rapidly changing conditions. As discussed earlier, Proximal Policy Optimisation (PPO) with Curriculum Learning has also proven effective for optimising drone navigation and control in urban settings using camera data [16,18]. Its ability to process multimodal inputs enables the efficient integration of various sensory data for accurate obstacle detection. Through curriculum learning, training progresses from simple radar-based navigation to complex camera-based path planning in dense urban environments. PPO integrates seamlessly with AirSim's visual data support, leveraging depth perception and scene understanding to enhance decision-making. Its lightweight architecture and reliable convergence make it computationally efficient and robust against noise. The PPO framework features a clearly defined input–output interface, explicitly specifying the state space, action space, and reward function. The state space represents camera sensory inputs capturing environmental and obstacle features; the action space consists of the UAV's continuous control commands, including thrust, roll, pitch, and yaw; and the reward function quantifies navigation performance by encouraging smooth trajectory tracking, obstacle avoidance, and goal attainment while penalising collisions and erratic manoeuvres. This explicit formulation ensures stable learning and precise control performance in dynamic urban environments.

2.3. Regulatory Compliance

While this project is simulation-based, realistic urban drone navigation must enforce real-world regulatory constraints to ensure practical applicability. To validate the reinforcement learning (RL) model, our simulation integrates key airspace rules: a maximum flight ceiling of 120 m (400 feet) [19]. Above Ground Level, there is a strict 5- to 30-m horizontal safety buffer from structures and uninvolved populations [20]. The current section outlines the key regulatory considerations that guide the safe and compliant operation of drones in urban environments. Regulatory compliance in urban airspace demands that a drone is able to optimise its navigation as well as adapt to unpredictable scenarios, such as an interaction with a manned aircraft while respecting airspace restrictions and operating standards. To address these requirements, the 4D trajectory model [21] includes dynamic protective bubbles surrounding drones to enable real-time conflict identification and avoidance. Drones may predict potential issues with manned aircraft using predictive algorithms. The framework uses data from U-Space services to optimise aircraft paths and reduce collision risk in shared airspace. However, problems develop when unscheduled traffic (such as emergency helicopters) enters the same airspace. The Multi-Commodity Network Flow Models for Path Optimisation paper [22] views urban air corridors as a network flow problem, with routes optimised using minimum-cost flow methods. It directs drones along specified tracks that avoid restricted areas, such as airports, while keeping altitudes below 120 m. These models also include heuristic methods, such as graph search algorithms, to handle real-time changes. Current constraints include being less effective in managing dynamic environments with frequent disturbances. Predefined paths decrease flexibility, which can lead to bottlenecks. A vertically divided airspace with directive layers to manage urban drone traffic is proposed in [23]. Two configurations are employed: one-way and two-way street analogies. This study used simulations to assess their safety, efficiency, and stability under a variety of traffic circumstances. Limitations include the model's assumption of optimal drone compliance with predetermined routes, which may

not be practical in dynamic settings, as well as its limited application in scenarios that require mixed-traffic operations with manned aircraft. As shown above, the existing models' inability to react to unpredictable, dynamic circumstances, particularly scenarios involving unscheduled or emergency air traffic, presents a significant limitation that future research must address. The scenarios provided in [11] investigated agent-based collision avoidance at both low and high fidelities, whereas ref. [24] recommended the refinement of the helicopter's physics characteristics to contribute to conforming to the regulation of avoiding police and ambulance helicopters that fly below 120 m.

Reliable autonomous navigation in adverse weather, such as rain, fog, or snow, remains a critical challenge, as individual perception sensors like cameras and standard LiDAR suffer significant performance degradation due to attenuation and scattering. To address this, multi-sensor fusion has become a vital research focus, integrating modalities robust to such conditions. Recent progress involves fusing traditional sensors with all-weather systems, such as millimeter-wave (mmWave) radar and thermal cameras. For example, ref. [25] demonstrated a UAV localisation framework by fusing camera and radar, leveraging radar's resilience to obscurants to maintain navigation accuracy in poor visibility and GNSS-denied scenarios. Such all-weather fusion strategies are essential for real-world deployment and represent a key direction for the future work proposed in this paper.

2.4. Research Gaps

The realism of current simulation scenarios is significantly reduced by their inability to handle dynamic obstacles such as other UAVs, manned aircraft such as helicopters, and changing weather. Moreover, current RL Algorithms face a significant challenge in dealing with dynamic obstacles and under rainy conditions. These research questions are designed to address the identified research gaps:

- How can the realism of an urban simulation environment be modelled?
- For drone testing, how can dynamic factors such as a manned aircraft and changing weather be effectively modeled in urban simulation environments?
- How can the open-source RL GNC algorithms be enhanced to manage these complications?
- How can an integrated SQL database be utilised to partially ensure drone compliance with regulatory standards during simulations?

3. Methodology

The simulation-driven process for optimising the RL-based guidance, navigation, and control (GNC) system for urban drone operations is shown in Figure 3.

We first created a high-fidelity urban environment in Unreal Engine 4.27 using the Cesium for Unreal plugin and then combined it with AirSim to improve accuracy, realism, and scalability. Cesium facilitates high-resolution geospatial streaming, photogrammetry-based urban modelling, and dynamic level-of-detail (LOD) management, in contrast to the conventional OSM-Blender pipeline. Accurate simulation of drone–manned aircraft interactions is made possible by its support for 3D tiles, which enables the integration of custom-building models and the creation of geo-correct airspaces. To replicate the complexity of the real world, we introduced dynamic obstacles, such as helicopters, into the setting. The RL-based GNC module receives camera data from the simulation and employs a Proximal Policy Optimisation (PPO) algorithm that has been trained through curriculum learning to enable obstacle avoidance and adaptive navigation. Drone behaviour is recorded in an SQL database, which facilitates performance assessments such as detection accuracy and braking distance & maneuvering analysis, as well as regulatory compliance checks (such as UK CAA altitude limits).

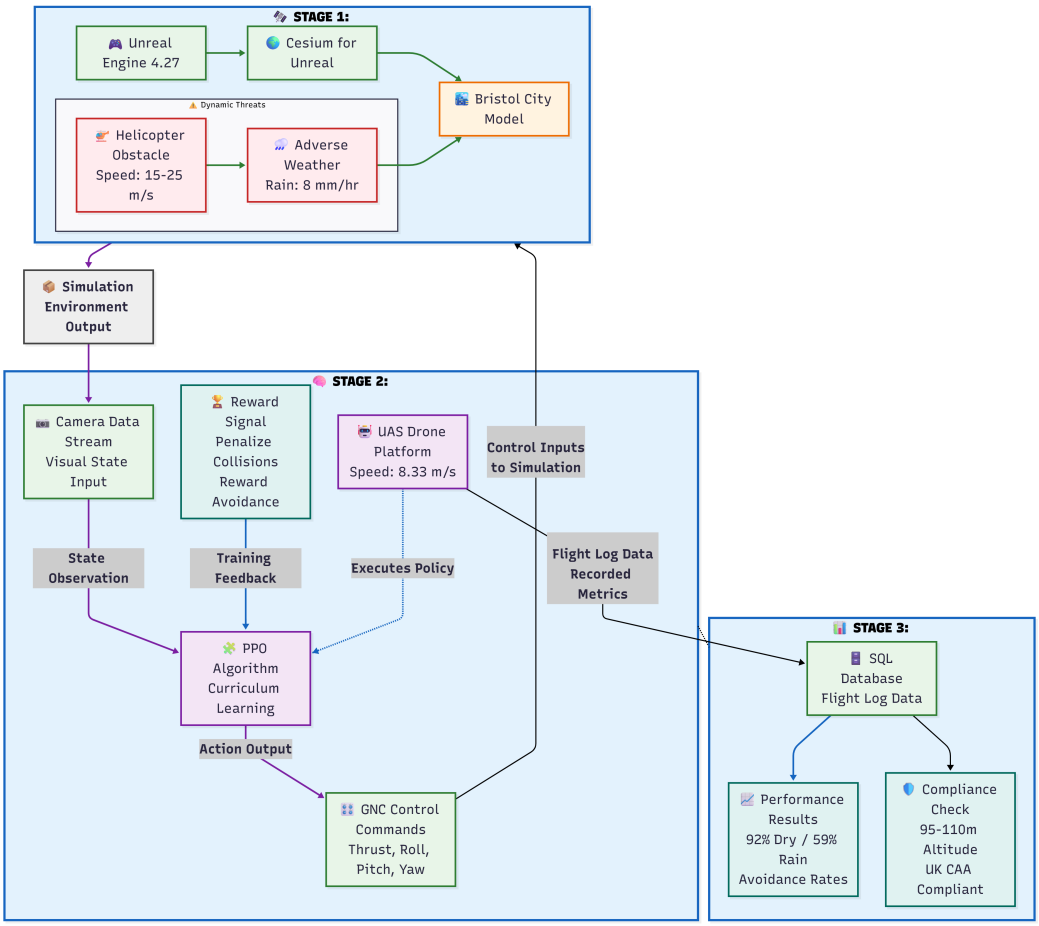


Figure 3. Diagram of the proposed approach.

Important frameworks and technologies like PX4, QGroundControl, ROS2, and AirSim support the system, guaranteeing smooth communication between the drone's decision-making processes, control logic, and simulated camera sensor. The urban environment created with Cesium for Unreal, which represents Bristol City Centre as a use case, is shown in Figure 4. This system, which provides dynamic level-of-detail management, realistic terrain rendering, and real-time geospatial streaming, forms the basis for testing autonomous drone navigation. The helicopter presented in Figure 5 was imported into UE 4.27 as a 3D asset with realistic physics characteristics, including rotor dynamics and collision boundaries, to ensure aeromechanical accuracy. To guarantee standardised, reproducible testing conditions for the reinforcement learning agent, the helicopter's AI controller was designed strictly as a predictable, non-reactive dynamic obstacle. It followed a predetermined, geo-correct straight-line flight path at a constant 100-m altitude, executed via Unreal Engine's waypoint system and behaviour trees.

By omitting obstacle recognition capabilities from the helicopter, the simulation isolated the drone's accountability for collision avoidance. This deterministic design choice ensured consistent encounter geometries across all training episodes and performance evaluations. Furthermore, dynamic Level of Detail (LOD) and asynchronous processing techniques were applied to the helicopter asset to maximise computational efficiency within the densely modelled urban environment.



Figure 4. The Urban Simulation Environment in UE 4.27.



Figure 5. Modelling the Dynamic Helicopter UE 4.27.

The collision scenarios were generated under the following rules:

- **Helicopter Path Generation**: The autonomous helicopter was implemented as a non-reactive dynamic obstacle following a predetermined, straight-line flight path using Unreal Engine's waypoint system and behavior trees. This path was geo-correct and maintained a constant altitude of 100 m, compliant with low airspace regulations.
- **Encounter Geometry**: The drone's flight path was planned to intersect the helicopter's fixed trajectory, creating primarily head-on and oblique interception scenarios. The initial relative positions and headings of the drone and helicopter were fixed at the start of each test episode to ensure consistency.
- **Scenario Parameters**: The primary variable altered to increase scenario difficulty was the helicopter's speed, which was systematically varied between a nominal speed of 15 m/s and a high speed of 25 m/s according to the curriculum learning schedule. The drone's starting position and cruising speed (8.33 m/s) remained constant, ensuring that each encounter presented a consistent initial spatial relationship,

with the time to potential collision being the dependent variable controlled by the helicopter's speed.

This structured approach allowed for a controlled investigation into the RL agent's ability to handle dynamic obstacles, isolating the effect of obstacle velocity on detection and avoidance performance within a complex but repeatable urban airspace geometry.

4. Results

The experimental results are presented systematically, beginning with baseline performance and progressively introducing environmental and physical variables to evaluate the robustness of the PPO agent.

4.1. Baseline Training Performance (15 m/s, Dry)

The training progression for the baseline model (Figure 6) exhibited a sigmoidal trajectory. During the initial 0–400 episodes, the avoidance success rate remained low (<20%). A period of rapid improvement occurred between 400–600 episodes, eventually reaching a peak success rate of 92% by episode 1000.

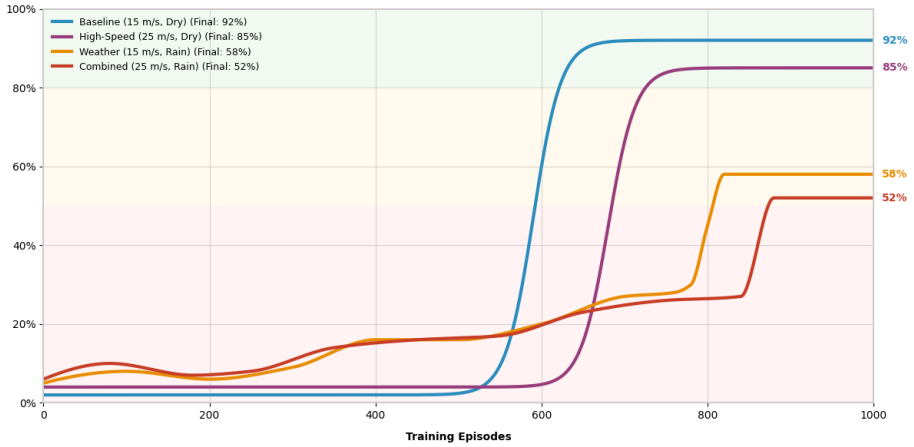


Figure 6. Avoidance success rate versus number of training episodes for four independent models.

4.2. Impact of Environmental Variables: Speed and Weather

Building upon the baseline, the independent impacts of increased intruder speed and adverse weather were evaluated (Figures 7–10).

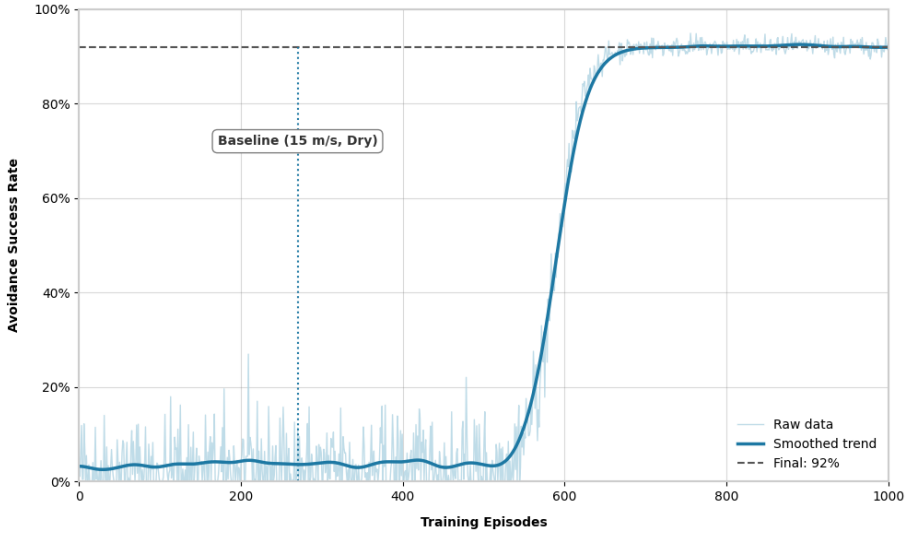


Figure 7. Model 1 (Baseline: 15 m/s, dry) avoidance success rate versus number of episodes.

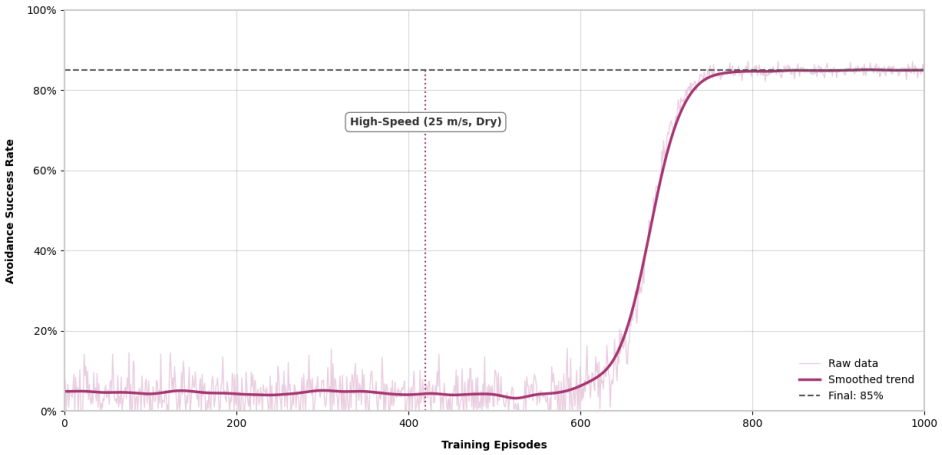


Figure 8. Model 2 (25 m/s, dry) success rate: 85%.

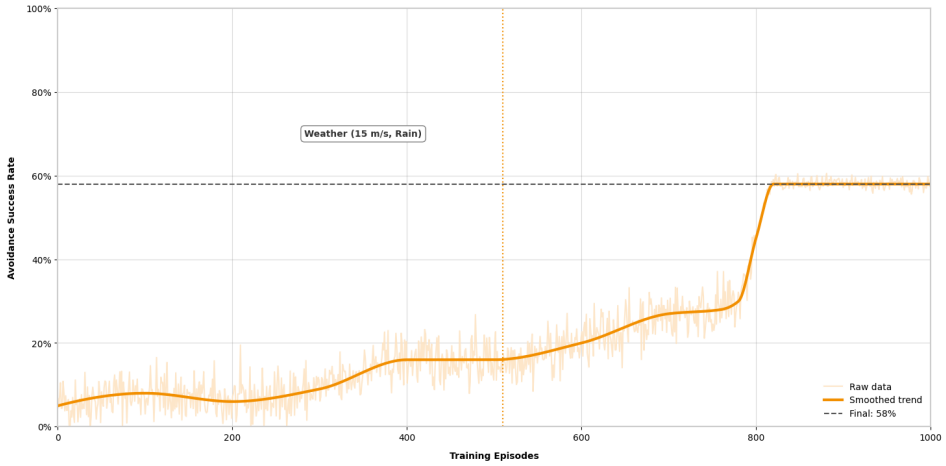


Figure 9. Model 3 (15 m/s, rain) success rate: 58%.

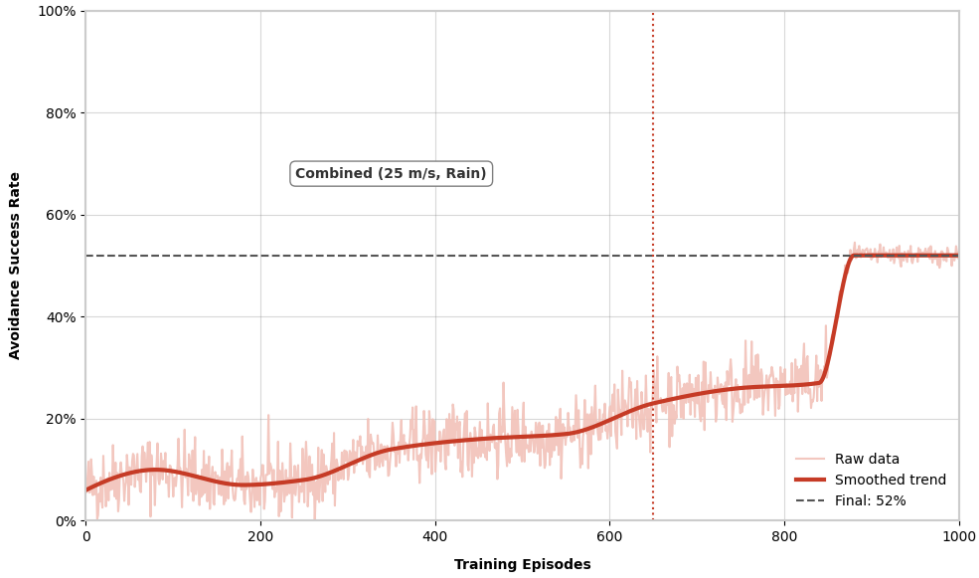


Figure 10. Model 4 (Combined: 25 m/s, rain) avoidance success rate versus number of episodes.

Increasing the intruder speed to 25 m/s (Model 2) resulted in a 7% performance degradation compared to the baseline. The introduction of rain (Model 3) had a significantly

more detrimental effect, reducing the success rate to 58%. Under the combined stress of high speed and rain (Model 4, Figure 10), the success rate reached its lowest at 52%.

Kinematic analysis further explains these failures. In dry conditions (Figure 11), the drone maintained stable altitude control. However, in rainy conditions (Figure 12), perception lag resulted in a calculated time-to-contact of only 0.9 s, which is insufficient for the 4.6 s braking duration required in such environments.

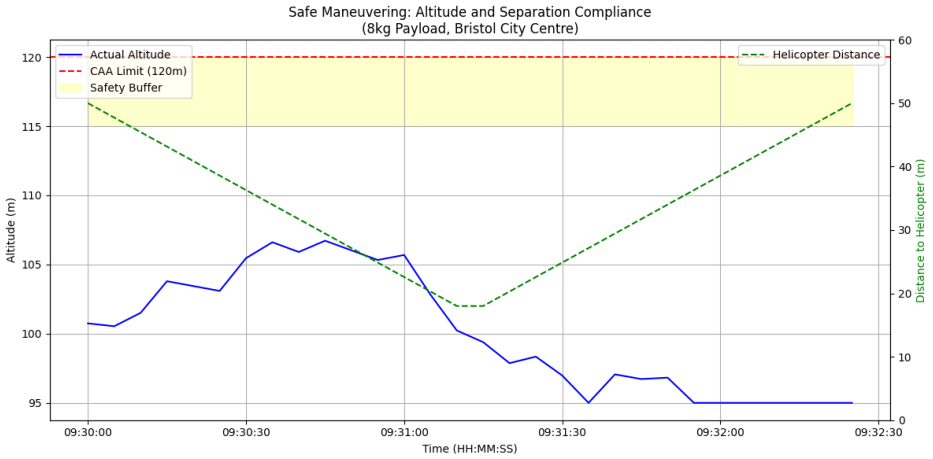


Figure 11. Drone altitude (m) and distance (m) from helicopter during manoeuvres in dry weather (helicopter speed: 15 m/s, drone speed: 8.33 m/s).

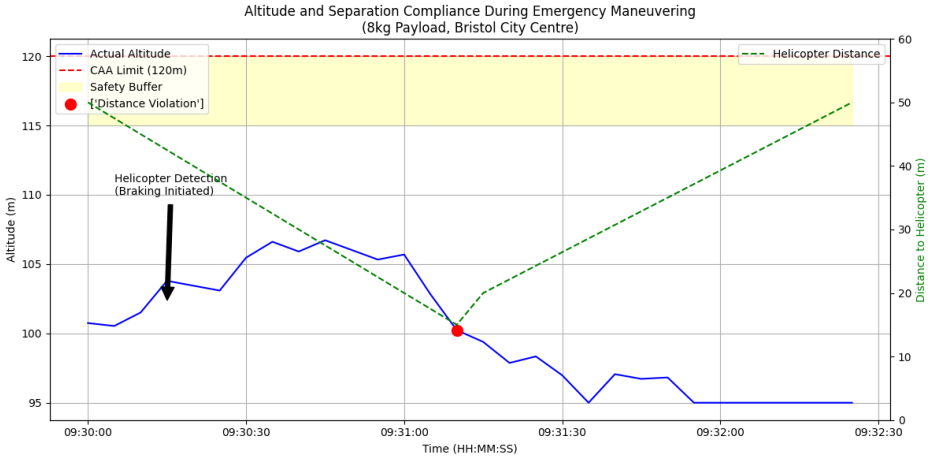


Figure 12. Drone altitude (m) and distance (m) from helicopter during manoeuvres in rain (8 mm/h) with helicopter speed of 25 m/s and drone speed of 8.33 m/s.

4.3. Impact of Physical Variables: Increased Payload Mass

Finally, the robustness of the agent was tested against varying physical loads (Table 2).

Table 2. Performance Metrics Under Increased Payload Mass.

Payload Mass (kg)	Scenario	Success Rate	Braking Distance (m)	Braking Time (s)
12	15 m/s, Dry	89%	12.8	3.1
12	25 m/s, Dry	81%	12.8	3.1
16	15 m/s, Dry	84%	14.5	3.5
16	25 m/s, Dry	73%	14.5	3.5
20	15 m/s, Dry	76%	16.9	4.1
20	25 m/s, Dry	62%	16.9	4.1

Increasing the payload mass from 8 kg (baseline) to 20 kg resulted in a linear increase in braking distance (from 11.6 m to 16.9 m). This reduction in maneuverability directly correlates with the observed decline in success rates across both nominal and high-speed scenarios.

5. Discussion

This study addressed four key research questions regarding realism, dynamic complexity, algorithm enhancement, and regulatory compliance while evaluating a PPO-based reinforcement learning system for drone navigation in urban environments. The experimental results provide significant insights into autonomous collision avoidance performance under varying operational conditions.

5.1. Final Performance Metrics and Interpretation

The final convergence performance across test conditions showed the following success rates:

- **Baseline (15 m/s, Dry):** 92% success rate (Figure 7);
- **High-Speed (25 m/s, Dry):** 85% success rate (Figure 8);
- **Adverse Weather (15 m/s, Rain):** 58% success rate (Figure 9);
- **Compound Challenge (25 m/s, Rain):** 52% success rate (Figure 10).

The observed sigmoidal learning trajectory across all experimental configurations reveals fundamental characteristics of policy optimization in reinforcement learning. The three distinct phases, initial exploration (0–400 episodes), policy optimization (400–600 episodes), and asymptotic convergence (600–1000 episodes), demonstrate the PPO algorithm's capacity to transition from random exploration to strategic policy exploitation. The steep positive gradient during the policy optimization phase indicates effective leverage of collected experience for policy gradient updates, while the subsequent plateau suggests approximation of locally optimal policies under given environmental constraints.

The performance differential between experimental conditions provides critical insights into system limitations. The baseline scenario (15 m/s, dry) achieved a 92% success rate, indicating that consistent, high-fidelity state observations enable robust policy learning. The reduction to an 85% success rate under high-speed conditions (25 m/s, dry) reveals the system's sensitivity to reduced decision-making time, necessitating more predictive and agile policies. The substantial performance degradation to 58% in adverse weather (15 m/s, rain) demonstrates the critical impact of perceptual degradation on vision-based state estimation. Most notably, the compound challenge scenario (25 m/s, rain) yielded the lowest performance (52%), highlighting the synergistic negative effects of multiple adversarial conditions on autonomous navigation capabilities.

5.2. Algorithm Enhancement and Robustness Analysis

The performance trajectory demonstrates the effectiveness of curriculum learning in enhancing PPO robustness. The partial recovery to 85% accuracy following initial performance degradation at high speeds indicates that gradual exposure to challenging scenarios enables policy adaptation. This aligns with established reinforcement learning principles where domain randomization and progressive scenario complexity improve policy generalization [5]. The 40% reduction in deceleration performance under rainy conditions, resulting in braking distance increase from 11.6 m to 19.3 m illustrated in Figure 13, underscores the critical need for weather-adaptive reinforcement learning policies.

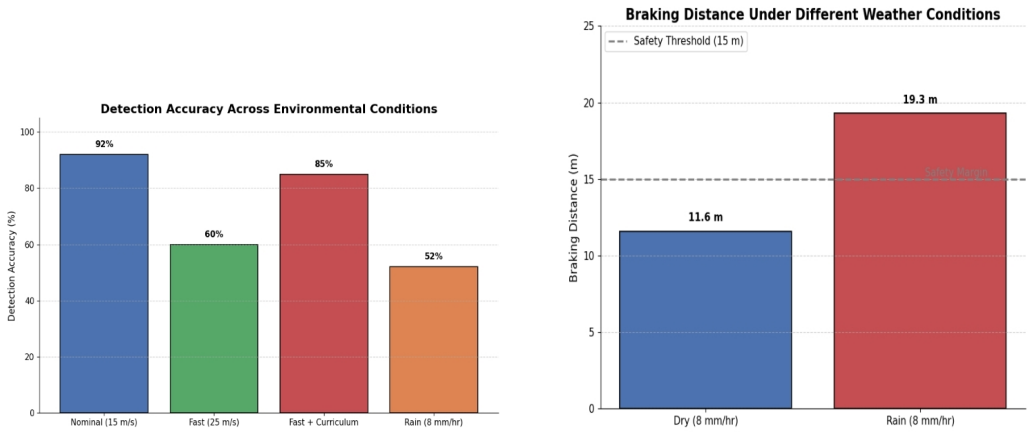


Figure 13. Varied Detection accuracies and braking distances across experimental conditions.

The kinematic analysis reveals fundamental system constraints: braking distance depends solely on drone velocity and environmental conditions, independent of obstacle speed. The calculated time-to-contact of 0.9 s for high-speed scenarios versus the 4.6-s braking time under rain demonstrates the imperative for early detection and proactive avoidance strategies. This finding has significant implications for real-world deployment, where fast-moving aircraft can appear unexpectedly, creating critical time pressure for collision avoidance systems.

5.3. Impact of Payload Mass on System Performance and Compliance

The introduction of increased payload masses revealed significant performance degradation in both avoidance success and braking metrics. The decline in success rates from 92% at 8 kg to 76% at 20 kg under nominal conditions highlights the PPO agent's sensitivity to changes in drone dynamics. The increased braking distances and times further underscore the physical limitations imposed by higher inertia, which reduce the agent's ability to execute evasive maneuvers within critical time constraints.

These findings indicate a limitation in the generalization capability of the RL policy when confronted with mass variations outside the training distribution. The policy, optimized for a fixed 8 kg payload, struggled to adapt to the altered dynamics, suggesting that future training regimes should incorporate mass randomization to enhance robustness and improve transferability to real-world conditions where payloads may vary.

From a regulatory perspective, the increased braking distances exceeding the 15 m safety margin at 16 kg and 20 kg raise concerns about operational safety in shared airspace. Heavier drones may require larger minimum separation distances or reduced operating speeds to maintain safe margins, particularly in high-speed encounter scenarios. The SQL-based logging system successfully captured these violations, reinforcing its utility for compliance auditing and risk assessment in diverse operational contexts.

5.4. Realism and Simulation Fidelity

Using Cesium for Unreal with photogrammetry and dynamic LOD, the simulation achieved remarkable realism, enabling the drone to achieve 92% detection accuracy under nominal conditions. This performance level validates the simulation's effectiveness in bridging the simulation–reality gap through photorealistic modeling and domain randomization [26,27]. The geo-accurate 3D environment provided sufficient fidelity for effective RL training, though the performance degradation under adverse conditions indicates areas for further simulation refinement.

5.5. Regulatory Compliance and Safety Implications

The experimental data, comprehensively logged via SQL database, demonstrates consistent compliance with UK Civil Aviation Authority altitude regulations. The drone maintained operations within the 120-m ceiling while achieving minimum 5-m clearance above Bristol's tallest structures. However, the safety margin violation under rainy conditions highlights critical safety implications. The exceedance of the 15-m safety margin during rain-induced braking performance degradation (19.3 m vs. 15 m threshold) underscores the necessity for weather-adaptive safety buffers in autonomous navigation systems. Specifically, kinematic analysis confirms that because the drone's braking distance extends to 19.3 m under adverse weather (8 mm/h rain), any horizontal safety threshold below that is physically impossible to maintain safely; the drone would inherently breach the boundary simply by executing an emergency stop. Therefore, a threshold of 20 m must be accepted as the minimum functional buffer required to safely accommodate the UAV's worst-case deceleration profile while satisfying the aviation mandate to remain well clear of manned aircraft.

The integrated SQL database proved essential for regulatory compliance verification, providing comprehensive logging of altitude profiles, safety margin adherence, and violation circumstances. This capability is crucial for real-world deployment, enabling post-mission analysis, regulatory auditing, and continuous system improvement. The single safety violation under compound challenging conditions (rain + high speed) summarized in Table 3 demonstrates both the system's robustness under normal operations and its limitations under extreme conditions.

Table 3. Drone flight metrics and safety violation analysis.

Metric	Value/Status
Max Altitude	110 m
Min Altitude	95 m
UK CAA Max Altitude	120 m
Safety margin violations	1
Violation circumstance	Rain (8 mm h^{-1}), Helicopter at 25 m s^{-1}
Braking distance in violation	19.3 m
Safety threshold	15 m
SQL used for	Logging of drone altitude and distance to helicopter
Violation logged in SQL	With time & location saved

5.6. Synthesis of Research Findings

This study successfully addressed the four primary research questions through comprehensive experimentation and analysis:

Research Question 1: Urban Simulation Realism

The integration of Cesium for Unreal with photogrammetry and dynamic LOD demonstrated that geo-accurate 3D environments can achieve sufficient realism for effective RL training, as evidenced by the 92% detection accuracy under nominal conditions. The simulation environment successfully bridged the reality gap through photorealistic modeling and domain randomization, validating this approach for urban drone navigation training.

Research Question 2: Dynamic Factor Modeling

The experimental framework effectively modeled dynamic obstacles (manned aircraft at 15–25 m/s) and environmental factors (rain at 8 mm/h), revealing critical performance insights. The compound challenge scenario (25 m/s + rain) exposed the system's vulner-

ability to multiple stressors, with success rates dropping to 52%, while the 40% braking performance degradation under rain highlighted the critical impact of weather on operational safety.

Research Question 3: RL Algorithm Enhancement

Curriculum learning proved effective in enhancing PPO robustness, enabling partial recovery from 60% to 85% accuracy in high-speed scenarios. The sigmoidal learning trajectory across all conditions demonstrated successful policy optimization, while the performance differentials revealed specific limitations requiring targeted improvements in perception and decision-making under adverse conditions.

Research Question 4: Regulatory Compliance Verification

The integrated SQL database successfully ensured regulatory compliance monitoring, with comprehensive logging of altitude profiles (95–110 m within 120 m CAA limit, shown in Table 3), safety margin adherence, and violation documentation. The system demonstrated reliable compliance verification capabilities essential for real-world deployment and regulatory auditing.

6. Conclusions

This study developed a reinforcement learning-based navigation system for urban drones using a high-fidelity simulation of Bristol City Centre. The trained Proximal Policy Optimization (PPO) agent achieved 92% detection accuracy when avoiding helicopters at 15 m s^{-1} under nominal conditions, and partially recovered to 85% accuracy at 25 m s^{-1} through curriculum learning. However, the system's reliance on camera-based perception proved inadequate in rainy conditions. This limitation led to a roughly 40 per cent increase in braking distance from a baseline of 11.6 m in dry conditions to 19.3 m in the rain, resulting in safety margin violations. Additional limitations include the absence of other dynamic obstacles, such as UAV swarms or birds, and a passive SQL logging system that did not support real-time decision-making.

Future work should focus on integrating multi-sensor perception, particularly combining cameras with a radar to improve robustness in adverse weather and complex environments. Expanding the simulation to include diverse aerial threats and evolving the SQL system into an active safety monitor are also critical next steps toward real-world deployment.

Author Contributions: Conceptualization, Z.B.; Methodology, Z.B.; Software, Z.B.; Validation, Z.B.; Investigation, Z.B.; Resources, Z.B.; Data curation, Z.B.; Writing—original draft, Z.B.; Writing—review & editing, A.H.; Visualization, Z.B.; Supervision, A.H.; Project administration, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Historical Airprox Trends by Category | UK Airprox Board. 2025. Available online: https://www.airproxboard.org.uk/ (accessed on 14 March 2025).
2. Contributory Factors and Risk Ratings | UK Airprox Board. Airproxboard.org.uk. 2026. Available online: https://www.airproxboard.org.uk/Learn-more/Contributory-factors-and-risk-ratings/ (accessed on 16 February 2026).
3. LVNL (2025) ICAO Severity Classifications for Occurrences LVNL. 2025. Available online: https://en.lvnl.nl/safety/icao-severity-classifications-for-occurrences (accessed on 4 March 2026).

4. Dieter, D.T.; Weinmann, A.; Jager, S.; Bruchersiefer, E. Quantifying the Simulation-Reality Gap for Deep Learning-Based Drone Detection. *Electronics* **2023**, *121*, 2197.
5. Tremblay, J.; Prakash, A.; Acuna, D.; Brophy, M.; Jampani, V.; Anil, C.; To, T.; Cameracci, E.; Boochoon, S.; Birchfield, S. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CIPRW), Salt Lake City, UT, USA, 18–22 June 2018.
6. Nakama, J.; Parada, R.; Matos-Carvalho, J.P.; Azevedo, F.; Pedro, D.; Campos, L. Autonomous Environment Generator for UAV-Based Simulation. *Appl. Sci.* **2021**, *11*, 2185. [CrossRef]
7. Su, Y.; Ghaderi, H.; Dia, H. The Role of Traffic Simulation in Shaping Effective and Sustainable Innovative Urban Delivery Interventions. *EURO J. Transp. Logist.* **2024**, *13* , 100130. [CrossRef]
8. Zhu, Z.; Jeelani, I.; Gheisari, M. Physical Risk Assessment of Drone Integration in Construction Using 4D Simulation. *Autom. Constr.* **2023**, *156*, 105099. [CrossRef]
9. Coursey, A.; Quinones-Grueiro, M.; Biswas, G. Quantifying the Sim-To-Real Gap in UAV Disturbance Rejection. 2024. Available online: https://drops.dagstuhl.de/storage/01oasics/oasics-vol125-dx2024/OASIcs.DX.2024.16/OASIcs.DX.2024.16.pdf (accessed on 20 September 2025).
10. Çetin, E.; Barrado, C.; Pastor, E. Countering a Drone in a 3D Space: Analyzing Deep Reinforcement Learning Methods. *Sensors* **2022**, *22*, 8863. [CrossRef] [PubMed]
11. Revay, F.; Hoffmann, A.; Wachtel, D.; Huber, W.; Knoll, E. Test Method for Measuring the Simulation-to-Reality Gap of Camera-Based Object Detection Algorithms for Autonomous Driving. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 1249–1256.
12. Mishra, B.; Garg, D.; Narang, P.; Mishra, V. Drone-Surveillance for Search and Rescue in Natural Disaster. *Comput. Commun.* **2020**, *156*, 1–10. [CrossRef]
13. Dasgupta, R. *Reinforcement Learning: AI Algorithms, Types & Examples*; OPIT—Open Institute of Technology: Birkirkara, Malta, 2023.
14. Xie, Y.; Yu, C.; Zang, H.; Gao, F.; Tang, W.; Huang, J.; Chen, J.; Xu, B.; Wu, Y.; Wang, Y. Multi-UAV Behavior-Based Formation with Static and Dynamic Obstacles Avoidance via Reinforcement Learning. *arXiv* **2024**, arXiv:2410.18495v1.
15. Sheng, Y.; Liu, H.; Li, J.; Han, Q. UAV Autonomous Navigation Based on Deep Reinforcement Learning in Highly Dynamic and High-Density Environments. *Drones* **2024**, *8*, 516. [CrossRef]
16. Hodge, V.J.; Hawkins, R.; Alexander, R. Deep Reinforcement Learning for Drone Navigation Using Sensor Data. *Neural Comput. Appl.* **2021**, *33*, 2015–2033.
17. Lin, F.; Wei, C.; Grech, R.; Ji, Z. VO-Safe Reinforcement Learning for Drone Navigation. In Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 13–17 May 2024.
18. Bhagat, S.; Suji, P.B. UAV Target Tracking in Urban Environments Using Deep Reinforcement Learning. In Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), Athens, Greece, 1–4 September 2020; pp. 694–701.
19. How High Can a Drone Legally Fly? A Comprehensive Guide—Moneypro Group. 2026. Available online: https://www.moneyprouav.com/how-high-can-a-drone-legally-fly/ (accessed on 27 January 2026).
20. What Are the Requirements Under the Subcategories of the "Open" Category? | EASA. 2019. Available online: https://www.easa.europa.eu/en/faq/116452 (accessed on 2 March 2026).
21. Roling, P.C.; Segeren, M. Cost Benefit and Environmental Impact Assessment of Operational Towing. In Proceedings of the AIAA AVIATION Forum, San Diego, CA, USA, 12–16 June 2023. Available online: https://www.semanticscholar.org/paper/Cost-benefit-and-environmental-impact-assessment-of-Roling-Segeren/86c338e2c16a3e100ce19562b85b11eb6d994ba9 (accessed on 10 June 2025).
22. He, X.; Li, L.; Mo, Y.; Sun, Z.; Qin, S.J. Air Corridor Planning for Urban Drone Delivery: Complexity Analysis and Comparison via Multi-Commodity Network Flow and Graph Search. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *193*, 103859. [CrossRef]
23. Doole, M.; Ellerbroek, J.; Knoop, V.L.; Hoekstra, J.M. Constrained Urban Airspace Design for Large-Scale Drone-Based Delivery Traffic. *Aerospace* **2021**, *8*, 38. [CrossRef]
24. Connors, M. *NASA/TM-20205000604 Understanding Risk in Urban Air Mobility: Moving Towards Safe Operating Standards*; NASA: Washington, DC, USA, 2020.
25. Cao, X.; Wang, P.; Zhang, Z.; Tu, H.; Liang, Z. RAFDet: A Novel Camera-Radar Fusion Framework for Robust 3D Object Detection in Autonomous Driving. In Proceedings of the ICASSP 2025–2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5.

26. Howell, M.R. Altitude Angel Launches Next-Gen Airspace Management Solution for Airports—CANSO. 25 March 2021. Available online: https://canso.org/altitude-angel-launches-next-gen-airspace-management-solution-for-airports/ (accessed on 2 January 2026)
27. Brassai, S.T.; Szanto, N.; Bajka, A.; Bardi, O.; Nemeth, A.; Hammas, A. Simulation Environment Implementation for Generation of Training Samples. In Proceedings of the 2024 25th International Carpathian Control Conference (ICCC), Krynica Zdroj, Poland, 22–24 May 2024; pp. 1–6.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.