A New Trajectory in UAV Safety: Leveraging Reinforcement Learning for Distance Maintenance Under Wind Variations

Xiaolin Xu M.S.

Jeffrey Sun

JAAER

A New Trajectory in UAV Safety: Leveraging Reinforcement Learning for Distance Maintenance Under Wind Variations

Xiaolin Xu[1a], **Jeffrey Sun**[2b]
[1]Purdue University, IN 47907 USA
[2]West Lafayette High School, IN 47906 USA
[a]syoushiroxu@outlook.com, [b]JeffreySun2@gmail.com

Abstract

In the field of aviation, safety is a critical cornerstone, and the operation of the Unmanned Aerial Vehicle (UAV) systems is deeply connected with this principle. A thorough analysis and rigorous simulation and testing of aircraft systems are essential to avoid severe safety hazards. This paper delves into the safety issue in UAV operations, specifically regarding maintaining minimum safety distances under fluctuating wind conditions. The study introduces a novel solution based on a Deep Deterministic Policy Gradient (DDPG) model, a reinforcement learning method. The DDPG model was trained using a simulated environment created through the Gazebo simulator, with values for wind and gust conditions derived from historical records at the KLAF airport at Purdue University. The model's performance was evaluated regarding maintaining safe distances under these conditions. The results indicate that the DDPG model can accurately predict safety distances with relatively low error rates when predicting under different weather conditions. The findings significantly contribute to UAV safety operations, suggesting the potential future utilization of reinforcement learning methods to study enhancing airspace efficiency and obstruction avoidance in UAVs.

Keywords: *Unmanned Aerial Vehicle, UAV, Safety Distance, Reinforcement Learning, Deep Deterministic Policy Gradient, Flight Safety, Airspace Efficiency, UAV Fleet Operation*

Introduction

To effectively mitigate safety risks, it is crucial to conduct thorough analyses and tests on UAV systems. Safety risks threaten lives, cargo integrity, and properties both onboard and along UAV flight paths. A critical aviation safety measure is maintaining a minimum safety distance, mandated by the Federal Aviation Administration (FAA) for all aircraft (Federal Aviation Administration, 2023). Ensuring UAVs keep safe distances from each other, and obstacles is essential, especially given the risks of collisions, which can damage property and endanger lives, particularly in urban areas. Environmental factors like wind add complexity, emphasizing the need for methods to maintain safe UAV separations. Despite the practice of setting safe UAV formations (Browne et al., 2022; Manathara & Ghose, 2011), this doesn't fully address safety distance maintenance in varying conditions, underlining the need for innovative solutions.

This study aims to fill this gap by introducing a novel approach based on reinforcement learning. While this machine learning approach does not guarantee absolute accuracy, it provides an estimated value with a low error rate. It offers a viable solution to the complex problem of maintaining minimal safety distances in UAV fleet operations. This study has the potential to unlock new possibilities in UAV formation control, offering an alternative way to prevent crashes and collisions. The approach could be implemented in UAV fleets that operate at lower altitudes, enabling them to navigate tunnel-like paths with more information and potentially easier path planning. This could significantly enhance airspace efficiency, such as shrinking the convoy size, so more UAVs could be fitted in the path, contributing to the broader goal of optimizing UAV operations.

The subsequent sections will delve into the specifics of the proposed approach, its implementation, and its potential impact on the future of UAV fleet management. First, the study's background is presented, providing context for the research question and its significance. The literature review evaluates existing research in the field, identifying gaps and justifying the need for the current study. Subsequently, the methodology is detailed, elucidating the techniques and processes used to gather and analyze data. The results highlight the findings and their potential implications. Then, the paper discusses ways to improve the current work. The paper concludes with a summary of the research, its contributions, limitations, and suggestions for future research.

Background

In this section, we will explore the multifaceted aspects of safety distance management in UAV fleet operations. We will delve into the critical role of wind and gust

conditions, the potential of machine learning approaches to addressing this issue, and the broader implications for airspace efficiency. In doing so, we aim to build upon the existing body of knowledge while simultaneously addressing the gaps this study seeks to fill.

The advent of UAVs has precipitated a paradigm shift in the aviation industry, ushering in an era of unprecedented possibilities for efficient airspace utilization. However, this technological revolution's promise is not without its challenges. One of the most pressing pertains to the assurance of safety in UAV fleet operations, particularly in the context of maintaining a minimum safety distance.

This domain's prevailing state of knowledge and practice is predominantly reactive. It focuses on collision avoidance strategies that necessitate course adjustments when a potential collision is imminent (Browne et al., 2022; Manathara & Ghose, 2011). While this approach has proven effective in scenarios characterized by relatively sparse airspace, it may not suffice as UAV operations become increasingly prevalent and the complexity of their interactions escalates.

The following work proposes a more proactive, anticipatory method to safety distance management in UAV fleet, particularly considering the dynamic and often unpredictable nature of wind conditions. We will then exam the impact of wind and gust conditions on UAV fleet operations, elucidating how these environmental factors can exert a profound influence on safety distance and discuss the inherent challenges in managing these effects. Subsequently, we will study the potential of machine learning approaches in predicting and managing safety distance under varying wind conditions. We will also discuss the broader implications for airspace efficiency, positing that a proactive, anticipatory approach to safety distance management can significantly enhance the utilization of airspace in UAV fleet operations.

The importance of studying minimum safety distance in UAV fleet operations, particularly in the context of changing wind conditions was addressed later. It highlights the need for innovative approaches, such as the machine learning method proposed in this study, to address this issue and enhance airspace efficiency. Despite the conspicuous absence of significant research in this area, this study aims to make a substantial contribution to the field by providing a novel perspective on safety distance management in UAV fleet operations.

Literature Review

The field of UAVs development, simulation, integration has seen significant advancements in recent years, particularly in the areas of deep learning with reinforcement learning (Cai et al., 2023; Cheng et al., 2021; Khairy et al., 2021), flight simulators (FlightGear, 2023; Gazebo, 2023; Microsoft Research, 2021), and machine learning-based minimum safety distance methods (Xu et al., 2023). This literature review aims to evaluate the existing body of knowledge in these areas, identify strengths and weaknesses, and argue for the necessity of the present study in filling a gap or building on a tradition.

Deep learning with reinforcement learning has emerged as a powerful approach in various domains, including UAV operations. Several deep learning methods have been explored, each with their unique strengths and limitations. For instance, Q-Learning, a model-free reinforcement learning algorithm, has been widely used due to its ability to learn from an environment without a model of the environment's dynamics. However, not all Q-Learning methods are suitable for all scenarios. Deep Q-Network (DQN) for example, does not support continuous action spaces (Arulkumaran et al., 2017), which are crucial for maintaining safety distance in UAV operations. Therefore, the Deep Deterministic Policy Gradient (DDPG) method, which supports continuous action spaces (Lillicrap et al., 2016), stands out as the most suitable option. DDPG, a model-free, off-policy algorithm, has proven effective in learning optimal policies in complex, continuous action spaces, making it an ideal choice for the present study.

Regarding flight simulators, numerous options are available, each offering unique capabilities within specific domains. FlightGear, for instance, excels in simulating various weather conditions, including rain and snow (FlightGear, 2023). AirSim, a multi-platform-based simulator, provides outstanding visualizations for vehicles and environments during simulation (Microsoft Research, 2021). However, these simulators lack the wide range of plugins and community support that Gazebo offers. For this study, the Gazebo simulator was chosen due to its realistic physics engine, support for multiple vehicles, customizable weather conditions, and the ability to import maps and modify models (Gazebo, 2023). Its compatibility with the Robot Operating System (ROS) (ROS, 2023) and the ease with which it can simulate varying wind conditions further underscore its suitability for this study (Mavlink, 2023). Despite these advantages, it is essential to note that Gazebo cannot simulate turbulence generated by other aircraft. This limitation must be considered when designing the training.

The literature on machine learning-based minimum safety distance methods for UAVs is sparse. The models for such computations are complex, and the industry norm has been to pre-set the formation or change the course before a collision occurs (Browne et al., 2022; Federal Aviation Administration, 2023; Manathara & Ghose, 2011). This approach, while effective, does not address the core issue of determining the minimum safety distance

under various wind conditions. The present study aims to fill this gap by focusing on this core issue.

In conclusion, by employing DDPG in a Gazebo simulation environment to address the problem of minimum safety distance in UAV operations, the present study seeks to fill a gap in the existing literature. By building on the strengths of deep learning with reinforcement learning and leveraging the capabilities of advanced flight simulators, this study aims to contribute to the field of UAV operations. This study's findings can enhance our understanding of safety distance management in UAV fleet operations and pave the way for future research in this rapidly evolving field.

Methodology

The primary objective is to devise a methodology capable of analyzing and maintaining the safety distance between fleet UAVs and proximate obstacles while flying. To address this research question, we employed a comprehensive and systematic methodology grounded in reinforcement learning, implemented through a series of designed simulation experiments.

The first step was to establish a robust and versatile simulation framework. This foundational stage was crucial in setting the stage for our subsequent experiments. To ensure the reliability and relevance of our framework, we referenced the structures used in previous works. This framework comprised three key components: the Gazebo simulation environment, the ROS communication network, and the PX4 autopilot flight control software. The Gazebo simulation environment provided a realistic and dynamic platform for our experiments. It allowed us to create a variety of scenarios and conditions, thereby enhancing the scope and applicability of our research. The ROS communication network facilitated efficient and reliable communication between different parts of our system, ensuring seamless integration and coordination. The PX4 autopilot flight control software, a highly advanced and widely used software in UAV research, provided reliable and precise control over the UAVs in our simulation (PX4, 2021). In this work, we will be focusing on the quad-rotors high quality and high-resolution Iris drone model.

The second phase of the methodology involved the creation of a DDPG training script. This script was instrumental in implementing the DDPG algorithm within the training loop. The script commenced with two identical models: the target and training networks. While similar at the onset, these networks played distinct roles in the training process. The target network essentially functioned as a delayed version of the training network. This delay was introduced to enhance the stability and efficiency of the training process. The rationale behind this approach is rooted in the minimization of the Mean Squared Bellman Error (MSBE), a key objective in reinforcement learning (Silver et al., 2014). By maintaining a delayed version of the training network, the target network provided a stable benchmark against which the training network's predictions could be compared. The target network was updated to match the training network whenever the latter underwent a significant update. This update was triggered when the training network exceeded a threshold defined by the Polyak averaging, a technique used to stabilize the learning process. Polyak averaging in the update process ensured that the target network remained a slightly outdated version of the training network, maintaining the stability of the training process. The process of updating the target network based on the Polyak averaging is shown as:

$$\phi_{\text{target}} \leftarrow \rho\phi_{\text{target}} + (1-\rho)\phi_{\text{train}},$$

where ρ is a hyperparameter from 0 to 1 and ϕ is the model network. We can see two models, both target and training, were contributed to the new target model. This will keep the experience learnt previously while adding new experience into the target model.

In the methodology employed for this study, the script was designed to feed the model with wind data generated randomly. This input data comprised five elements: wind speed, gust speed, and the force vectors in the x, y, and z-axis. To ensure compatibility with the model, these values were scaled down to hyperparameters ranging from 0 to 1. This normalization process facilitated the data processing by the model, enhancing the efficiency and accuracy of the predictions from previous work. Then the target network evaluated the inputs and generated predictions. These predictions consisted of three hyperparameter outputs ranging from 0 to 1. These outputs represented the maximum displacement of the UAV's origin and the maximum displacement towards the left and right of the UAV. These displacement predictions were based on the maximum action Q values in the training network, a vital component of the reinforcement learning process. These predictions were carried out using the minimization of MSBE loss with stochastic gradient descent. This function, a variant of the Mean Squared Error (MSE) used in regression analysis, quantifies the average squared difference between the estimated and actual values in the context of the Bellman equation, a fundamental equation in reinforcement learning. The use of the MSBE function in the evaluation process is illustrated below. It first retrieved the Q values from the future state, applied a discount and learning rate to generate the expected return. Then it used this expected return to merge with the training model for learning. For instance,

$$L(\emptyset, \mathcal{D}) = E_{(s,a,r,s',d)\sim\mathcal{D}}[\left(Q_{\emptyset_{train}}(s,a) - \left(r + \gamma(1-d)Q_{\emptyset_{target}}\left(s', \mu\theta_{target}(s')\right)\right)\right)^2]$$

where $\mu_{\theta_{\text{target}}}$ is the target policy, $Q_{\emptyset}(s,a)$ is the action-value function to output the results, s is the state,

s' is the next state, a is the action, D is a set of transitions (s,a,r,s',d), in which d indicates whether s' is terminal (true means 1 and false means 0), r is the reward, γ is the discount factor, and finally, E is the expected return. The script then initiated the simulation using the randomly generated wind conditions. This process facilitated the observation of the UAV's behavior within the simulator. The UAV's movements, along with other pertinent details necessary for evaluating the accuracy of the prediction, were meticulously recorded. This data collection was crucial in generating a reward value for the algorithm, a key component of the reinforcement learning process.

The evaluation process was straightforward, employing the MSE as the metric for comparing the predicted data with the actual measurements. The MSE, a widely used measure in regression analysis, quantifies the average squared difference between the estimated and actual values, providing a robust measure of prediction accuracy. Upon the generation of the reward, it was subsequently added to the script buffer. This buffer served as a repository for the collected data, storing it until a sufficient amount had been accumulated to commence the training of the network. This approach ensured that the training was based on a comprehensive dataset, enhancing the robustness of the learning process. The training of the network was based on policy learning, a method well-suited to the DDPG model used in this study. Given that DDPG supports continuous action spaces, it was possible to employ gradient ascent in the training process. This optimization technique, which seeks to maximize a function by iteratively moving in the direction of steepest ascent, was instrumental in refining the model's predictions. The use of gradient ascent in the training process is:

$$\max_{\theta} E_{s\sim\mathcal{D}}\left[Q_{\phi_{train}}\left(s, \mu\theta_{train}(s')\right)\right]$$

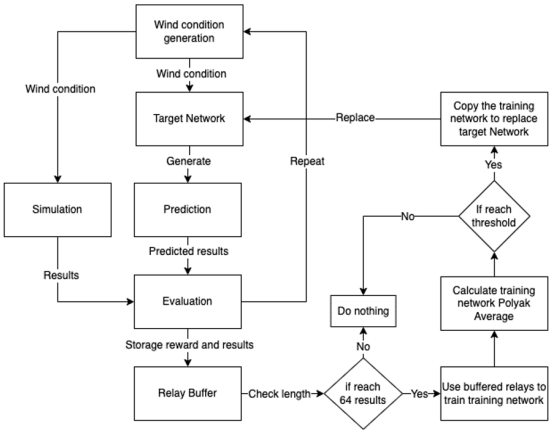
where $\mu_{\theta_{\text{train}}}$ is the training policy, $Q_{\phi}(s,a)$ is the action-value function to output the results, s is state and s' is next state, E is the expected return.

While reinforcement learning has been successfully applied in various domains using Q-learning, the need for a compact model that can operate on a drone necessitated the use of a deep network. Given that the safety distance should be treated as a continuous action space, we opted for DDPG over DQN. This choice was further justified by the need to incorporate the distance to obstacles into the training process, which would enhance vehicle passing ability in tunnel-like environments. The entire workflow was demonstrated in Figure 1.

The third step involved enhancing the realism and complexity of our simulation environment. We modified the environment to create a tunnel-like environment with four walls. This setup was intended to mimic the challenges UAVs might encounter in real-world operations, such as navigating confined spaces or avoiding obstacles.

Figure 1

The Workflow of the Integrated System



In the fourth step, we equipped the drone with a lidar sensor for our simulation. This sensor measured the distance from the UAV to each wall. By incorporating this sensor data into our DDPG training script, we trained our model to maintain a safe distance from the walls, effectively simulating the task of avoiding obstacles in flight.

Our methodology combines advanced simulation techniques, reinforcement learning algorithms, and sensor data to address the problem of safety distance management in UAV fleet operations. Each step contributes to answering our research question, providing a comprehensive and realistic exploration of this issue and a solid platform to conduct our research.

Results

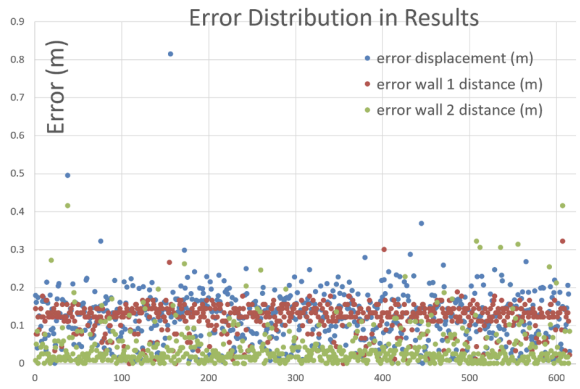
An extensive series of tests were meticulously conducted to devise a methodology capable of analyzing and maintaining the safety distance between UAVs in a fleet and proximate obstacle while in transit. The data for these tests were derived from the historical records of the National Oceanic and Atmospheric Administration (NOAA) at the KLAF airport (Aviation Weather Center, 2023). This historical weather data, encompassing precise historical wind and gust information, was utilized to recreate the weather conditions within the Gazebo simulator. This facilitated testing the accuracy of the trained model using the Iris drone model.

The model used for this study was a DDPG model with 1k parameters. This model was trained for 10k epochs using simulated wind and gust data. The model has three outputs: maximum shifting displacement, maximum shifting displacement to the left (denoted as wall

1), and maximum shifting displacement to the right (denoted as wall 2). The training process was designed to optimize the model's ability to predict the safety distance under varying wind conditions, enhancing its applicability to real-world UAV operations. 10k epochs will result in different time consumption due to various equipment for other users, but in our settings, the total training time for this work was approximately 79 hours. For each prediction, which includes all three values, the processing time costs an average of 0.7 milliseconds.

Figure 2

Error Distribution in Verification



The tests were designed with a specific constraint: the maximum displacement was restricted to 1.5 meters or less. This constraint was imposed to ensure the relevance and applicability of the results to real-world UAV operations, where maintaining a minimum safety distance is paramount. The respective errors for displacement prediction were meticulously recorded for wall 1 and wall 2. The analysis of the results, as demonstrated in Figure 2, revealed some interesting findings. The average error in predicting displacement for wall 1, wall 2, and overall displacement was 11.95 cm, 3.90 cm, and 12.39 cm, respectively. These figures represent the average deviation from the predicted values, providing a quantitative measure of the accuracy of the trained model. To ensure the highest level of safety and account for 99% of the error, results suggest that the flight control system should incorporate a safety buffer to the prediction in displacement, Wall 1, and Wall 2 when considering control, avoidance, and path planning. Specifically, safety buffers of 12.08 cm, 11.85 cm, and 3.56 cm should be added to the predictions for displacement, wall 1 and wall 2, respectively. By maintaining the suggested safety buffers above in the path planning algorithm, the vehicles can mitigate 99% of the risk in a mid-air collision and ensure the safe operation of the UAVs under varying wind conditions.

Discussion

The accuracy analysis of the trained DDPG model provided insights into the safety distance of UAV operations under varying wind conditions. The average error rates in the prediction demonstrated above were 8.269%, 7.971%, and 2.597% for displacement from waypoints and distance to Wall 1 and 2, respectively. These figures indicate that the model can provide a reliable estimate of the drone's reaction to the impulse of gust forces, thereby enhancing the safety and efficiency of UAV fleet operations. The Gazebo simulator also contains other high-quality UAV models, such as Typhoon HX480 and fixed-wing aircraft. Researchers can adopt and use these models to conduct their research or easily verify similar work. However, since the aerodynamics for different models are drastically different, the DDPG model trained on, for example, the Iris model, cannot be used on the Typhoon HX480 UAV model.

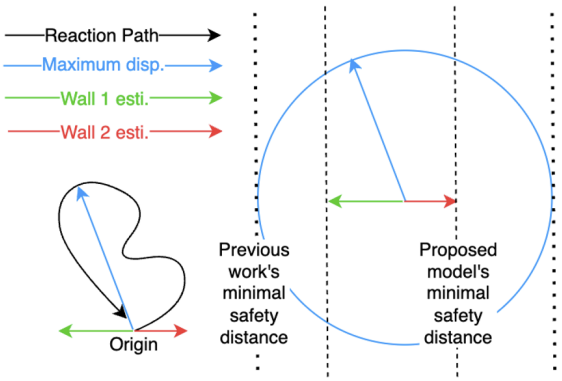
The accuracy of the DDPG model in predicting the safety distance under varying wind conditions has significant implications for real-world UAV operations. The more information the flight control system has, the better it can ensure safety. Therefore, understanding the minimum safety distance between a UAV and its surrounding obstacles is crucial in avoiding collisions. This information can enhance the safety of UAV operations, particularly when UAVs are traveling in a fleet or urban area. The model can help make the airspace more efficient by accurately predicting the safety distance. This is particularly relevant in congested airspaces where the ability to predict and maintain safety distances accurately can significantly reduce the risk of collisions and improve overall operational efficiency.

The model-based DDPG's suitability for path planning in tunnel-like environments is another significant finding of this study. This model is easy to train and deploy on real-world UAVs to handle wind and gust conditions. Its small size, and the ease with which it can be trained using only simulation data, makes it particularly suitable for this task. The path planning algorithm can enable the drone to travel in tighter tunnels by accurately estimating the distance shifted from the left or right. This represents a significant improvement over previous models, which only considered displacement, demonstrated in Figure 3. The ability to accurately predict shifts in position due to wind and gust forces can enable more precise path planning, allowing UAVs to navigate through tighter spaces and avoid obstacles more effectively.

In addition to enhancing safety, the DDPG model could also improve the efficiency of airspace usage. The model could enable a fleet of UAVs to maintain a dynamic formation based on wind and gust conditions. This allows for more efficient use of airspace, as the UAVs may adjust

Figure 3

Visualization of Proposed Work Improvements Compared to Previous Work



their formation in response to changing conditions. This dynamic formation adjustment capability could lead to more efficient airspace usage, allowing for closer spacing of UAVs in a fleet without compromising safety. This is particularly beneficial in congested airspaces, where efficient use of space is crucial.

Future Work

While the Gazebo simulator was chosen for its realistic physics engine and compatibility with ROS, it cannot simulate turbulence generated by other aircraft. This limitation could potentially impact the accuracy of the model's predictions in real-world scenarios where such turbulence is present. Additionally, while beneficial for testing the model, the study's reliance on historical data for recreating weather conditions may not fully capture the variability and unpredictability of real-world wind conditions.

Future research could aim to address these limitations. For instance, incorporating turbulence generated by other aircraft into the simulation could enhance the model's accuracy. Additionally, real-time wind data could further improve the model's predictions. Future studies could also explore applying the DDPG model in other areas of UAV operations, such as dynamic formation adjustment for more efficient airspace usage. Furthermore, the potential of the DDPG model in path planning in tunnel-like environments could be further explored, potentially leading to more precise path planning and improved obstacle avoidance.

Conclusion

The presented study holds significant implications for the field of UAVs, particularly in the context of safety distance prediction under varying wind conditions. The research employed a DDPG model, demonstrating its efficacy in maintaining safety distances between UAVs and proximate obstacles. The model's accuracy, as evidenced by the relatively low error rates, underscores its potential for real-world applications, contributing to the safety and efficiency of UAV operations.

One of this study's key strengths is its innovative use of the DDPG model, which supports continuous action spaces, making it particularly suitable for this task. The model's adaptability and ease of training using only simulation data further enhance its applicability. Moreover, the study's rigorous testing approach, involving the recreation of weather conditions within the Gazebo simulator using historical data, adds to the robustness of the findings.

The study demonstrated the DDPG model's potential in predicting safety distances under varying wind conditions by establishing the system design and model usage in a tunnel-like urban environment. Despite its limitations, the study opens new avenues for future research, paving the way for safer and more efficient UAV operations.

References

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, *34*(6), 26–38. https://doi.org/10.1109/MSP.2017.2743240

Aviation Weather Center. (2023). *AWC - Meteorological Aerodrome Reports (METARs)*. https://aviationweather.gov/data/metar/

Browne, J. P., Neuhart, C., Moleski, T. W., & Wilhelm, J. (2022). Minimal deviation from mission path after automated collision avoidance for small fixed wing uavs. *AIAA SCITECH 2022 Forum*. https://doi.org/10.2514/6.2022-0275

Cai, M., Fan, S., Xiao, G., & Hu, K. (2023). Deep reinforcement learning-based uav path planning algorithm in agricultural time-constrained data collection. *Advances in Electrical and Computer Engineering*, *23*(2), 101–108. https://doi.org/10.4316/AECE.2023.02012

Cheng, Z., Shen, H., Wang, Y., Wang, M., & Bai, G. (2021). Deep reinforcement learning based uav assisted svc video multicast. *Ji Suan Ji Ke Xue*, *48*(9), 271–277.

Federal Aviation Administration. (2023). *Air traffic control (jo 7110.65aa) [order]*. https://www.faa.gov/

regulations_policies/orders_notices/index.cfm/go/document.current/documentnumber/7110.65

FlightGear. (2023). *Features*. https://www.flightgear.org/about/features/

Gazebo. (2023). *Features and benefits*. https://gazebosim.org/features

Khairy, S., Balaprakash, P., Cai, L. X., & Cheng, Y. (2021). Constrained deep reinforcement learning for energy sustainable multi-uav based random access iot networks with noma. *IEEE Journal on Selected Areas in Communications*, *39*(4), 1101–1115. https://doi.org/10.1109/jsac.2020.3018804

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*. https://doi.org/10.48550/arXiv.1509.02971

Manathara, J. G., & Ghose, D. (2011). Reactive collision avoidance of multiple realistic uavs. *Aircraft Engineering and Aerospace Technology*, *83*(6), 388–396. https://doi.org/10.1108/00022661111173261

Mavlink. (2023). *Px4/px4-sitl_gazebo-classic: Set of plugins, models and worlds to use with osrf gazebo simulator in sitl and hitl*. https://github.com/PX4/PX4-SITL_gazebo-classic

Microsoft Research. (2021). *Welcome to airsim*. https://microsoft.github.io/AirSim/

PX4. (2021). *Software overview: What is px4?* https://px4.io/software/software-overview/

ROS. (2023). *ROS Noetic Ninjemys*. http://wiki.ros.org/noetic

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., & Riedmiller, M. (2014). Deterministic policy gradient algorithms. *Proceedings of the 31st International Conference on Machine Learning, in Proceedings of Machine Learning Research*, *32*(1), 387–395. https://proceedings.mlr.press/v32/silver14.html

Xu, X., Sun, J., & Hu, H. (2023). Simulator based mission optimization for swarm uavs with minimum safety distance between neighbors. *AIAA AVIATION 2023 Forum*. https://doi.org/10.2514/6.2023-4453