Article
文章

Adaptive Collision Avoidance for Multiple UAVs in Urban Environments

城市环境中多无人机自适应避障

Jinpeng Zhang [1,2] , Honghai Zhang [1,2,*] , Jinlun Zhou [1,2] , Mingzhuang Hua [2,3] , Gang Zhong [1,2] . Jang Zhong [1,2] . and Hao Liu [2,4]

张津鹏 [1,2] ，张洪海 [1,2,*] ，周金伦 [1,2] ，华明庄 [2,3] ，钟刚 [1,2] 。钟江 [1,2] 和刘浩 [2,4]

1 College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; zjp0401@nuaa.edu.cn (J.Z.); jinlunzhou@nuaa.edu.cn (J.Z.); zg1991@nuaa.edu.cn (G.Z.)

1 南京航空航天大学民航学院，南京 211106，中国；zjp0401@nuaa.edu.cn (J.Z.)；jinlunzhou@nuaa.edu.cn (J.Z.)；zg1991@nuaa.edu.cn (G.Z.)

2 National Key Laboratory of Air Traffic Flow Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; huamingzhuang@nuaa.edu.cn (M.H.); hliu@nuaa.edu.cn (H.L.)

2 南京航空航天大学空中交通流量管理国家重点实验室，南京 211106，中国；huamingzhuang@nuaa.edu.cn (M.H.)；hliu@nuaa.edu.cn (H.L.)

3 College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics, Liyang 213300, China

3 南京航空航天大学通用航空与飞行学院，江苏省溧阳 213300，中国

4 College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

4 南京航空航天大学理学院，南京 211106，中国

* Correspondence: honghaizhang@nuaa.edu.cn; Tel.: +86-13914766151

* 联系人:honghaizhang@nuaa.edu.cn；电话:+86-13914766151

Abstract: The increasing number of unmanned aerial vehicles (UAVs) in low-altitude airspace is seriously threatening the safety of the urban environment. This paper proposes an adaptive collision avoidance method for multiple UAVs (mUAVs), aiming to provide a safe guidance for UAVs at risk of collision. The proposed method is formulated as a two-layer resolution framework with the considerations of speed adjustment and rerouting strategies. The first layer is established as a deep reinforcement learning (DRL) model with a continuous state space and action space that adaptively selects the most suitable resolution strategy for UAV pairs. The second layer is developed as a collaborative mUAV collision avoidance model, which combines a three-dimensional conflict detection and conflict resolution pool to perform resolution. To train the DRL model, in this paper, a deep deterministic policy gradient (DDPG) algorithm is introduced and improved upon. The results demonstrate that the average time required to calculate a strategy is 0.096 s , the success rate reaches 95.03% , and the extra flight distance is 26.8 m , which meets the real-time requirements and provides a reliable reference for human intervention. The proposed method can adapt to various scenarios, e.g., different numbers and positions of UAVs, with interference from random factors. The improved DDPG algorithm can also significantly improve convergence speed and save training time.

摘要: 低空空域中无人驾驶飞行器 (UAVs) 数量的增加严重威胁到城市环境的安全。本文提出了一种针对多无人飞行器 (mUAVs) 的自适应避撞方法，旨在为存在碰撞风险的 UAVs 提供安全指导。提出的方法被构建为一个双层决策框架，考虑了速度调整和重新规划策略。第一层建立为一个具有连续状态空间和动作空间的深度强化学习 (DRL) 模型，该模型能够自适应地为 UAV 对选择最合适的解决策略。第二层开发为一个协作 mUAV 避撞模型，该模型结合了三维冲突检测和冲突解决池来执行解决。为了训练 DRL 模型，本文引入并改进了深度确定性策略梯度 (DDPG) 算法。结果显示，计算策略所需的平均时间为 0.096 s ，成功率达到 95.03% ，额外飞行距离为 26.8 m ，满足了实时要求，并为人类干预提供了可靠的参考。提出的方法能够适应各种场景，例如，不同数量和位置的 UAVs，以及随机因素的干扰。改进的 DDPG 算法还可以显著提高收敛速度并节省训练时间。

check for updates
检查更新

Citation: Zhang, J.; Zhang, H.; Zhou, J.; Hua, M.; Zhong, G.; Liu, H. Adaptive Collision Avoidance for Multiple UAVs in Urban Environments. Drones 2023, 7, 491.

引用: 张杰；张华；周健；华明；钟国；刘华。城市环境中多无人飞行器的自适应避撞。无人机 2023，7，491。https://doi.org/10.3390/drones7080491
无人机 7080491
Academic Editor: Pablo Rodríguez-Gonzálvez
学术编辑:Pablo Rodríguez-Gonzálvez
Received: 5 July 2023
收到日期:2023 年 7 月 5 日
Revised: 20 July 2023
修订日期:2023 年 7 月 20 日
Accepted: 23 July 2023
接受日期:2023 年 7 月 23 日
Published: 27 July 2023 creativecommons.org/licenses/by/ 4.0/).
发表日期:2023 年 7 月 27 日 creativecommons.org/licenses/by/ 4.0/)。

Keywords: urban air traffic management; collision risk; adaptive collision avoidance; deep reinforcement learning; multiple unmanned aerial vehicles
关键词: 城市低空交通管理；碰撞风险；自适应避障；深度强化学习；多无人航空器

1. Introduction

1. 引言

Urban low-altitude airspace is an important natural resource that possesses great socioeconomic value, and rational management of urban airspace is of great significance in alleviating traffic congestion and reducing the rate of ground traffic accidents [1, 2] . As the main subjects of urban air traffic, unmanned aerial vehicles (UAVs) have attracted widespread attention due to their flexibility, convenience, and low cost. By the end of 2022, China had 700,000 registered UAV owners, 15,130 companies operating UAVs, 958,000 registered UAVs, and about 57,000 h of average daily flight [3]. It can be foreseen that with the development of urban air traffic, the types of tasks performed by UAVs will inevitably tend to diversify, showing great promise in fields such as tourism, rescue, and logistics, and with this comes an increase in urban air traffic flow and the complication of flight situations.

城市低空空域是一种具有重要社会经济价值的自然资源，合理管理城市空域对于缓解交通拥堵和降低地面交通事故率具有重要意义 [1, 2] 。作为城市空中交通的主要参与者，无人机 (UAVs) 因其灵活性、便利性和低成本而受到广泛关注。截至 2022 年底，中国拥有 70 万注册无人机所有者，15130 家公司运营无人机，注册无人机数量达到 95.8 万台，平均每日飞行量约为 57,000 h [3]。可以预见，随着城市空中交通的发展，无人机执行的任务类型将不可避免地趋向多样化，在旅游、救援和物流等领域展现出巨大潜力，随之而来的将是城市空中交通流量的增加和飞行情况的复杂化。

As the number and size of UAVs increases, their operation in urban airspace will present additional security threats. The dense distribution of buildings, the complex structure of the airspace, and the high density of aircraft make it extremely easy for accidents such as dangerous approaches or even collisions to occur. Therefore, in the face of limited airspace resources, means of effectively avoiding risk of collision have become a primary issue that needs to be addressed in order to build urban air traffic demonstration areas and develop the low-altitude economy. However, the traditional collision avoidance algorithms lack sufficient success rates, and do not satisfy the safety interval criteria and real-time requirements in multi-target and high-density urban scenarios. In addition, these methods generate collision avoidance trajectories based on discrete state space, and the selectable actions are also discrete, and are not able to adequately reflect the flexibility of UAVs.

随着无人机数量和体积的增加，其在城市空域的运行将带来额外的安全威胁。建筑物的密集分布、空域结构的复杂性和飞机的高密度使得危险接近甚至碰撞等事故极易发生。因此，面对有限的空域资源，有效避免碰撞风险的手段已成为构建城市空中交通示范区和开发低空经济的首要解决的问题。然而，传统的碰撞避免算法在多目标和高度城市场景中的成功率不足，且不能满足安全间隔准则和实时性要求。此外，这些方法基于离散状态空间生成避障轨迹，可选择的行为也是离散的，不能充分反映无人机的灵活性。

To address these problems, we propose an innovative two-layer resolution framework for mUAVs based on DRL, which can adaptively provide avoidance strategies for UAVs based on continue action space and ensure that each UAV has decision-making capability, thus significantly improving the success rate and computational efficiency.

为了解决这些问题，我们提出了基于深度强化学习 (DRL) 的创新两层解决框架，用于微型无人机 (mUAVs)，该框架能够基于连续动作空间自适应地为无人机提供避障策略，并确保每架无人机都具有决策能力，从而显著提高成功率并提升计算效率。

1.1. Related Prior Work

1.1. 相关前期工作

Many studies have been performed proposing methods for UAV collision avoidance. In general, existing methods can be grouped into the following three categories: heuristic optimization methods, optimal control theory methods, and artificial intelligence methods.

许多研究已经提出了用于无人机避障的方法。总的来说，现有方法可以分为以下三类: 启发式优化方法、最优控制理论方法和人工智能方法。

(1) Heuristic optimization methods

(1) 启发式优化方法

Heuristic optimization methods divide the conflict process into a series of discrete state spaces and then perform an optimal search for approximate solutions in a certain cooperative manner [4], and primarily include the swarm intelligence optimization method [5] $\mathrm{A}^*, \mathrm{D}^*$. Zeng et al. combined the ant colony algorithms and the A* algorithm to solve the unmanned ground vehicle (UGV) scheduling planning problem, avoiding conflicts during simultaneous path planning for UGVs at a lower cost [6]. Zhao et al. considered collision probability and the intention information of intruders, using the A* algorithm to optimize trajectory planning to avoid collision risks [7]. Yun et al. applied the enhanced D* lite algorithm in the field of robot path navigation in unknown dynamic environments [8]. Furthermore, these methods are usually combined with other algorithms to solve the conflict problem, such as clustering methods [9] and Legendre Pseudo spectral Method [10].

启发式优化方法将冲突过程划分为一系列离散状态空间，然后在某种合作方式下进行最优搜索以寻找近似解 [4]，主要包括群体智能优化方法 [5] $\mathrm{A}^*, \mathrm{D}^*$ 。曾等人结合了蚁群算法和 A* 算法来解决无人地面车辆 (UGV) 调度规划问题，在同时进行路径规划时以较低成本避免冲突 [6]。赵等人考虑了碰撞概率和入侵者的意图信息，使用 A* 算法优化轨迹规划以避免碰撞风险 [7]。云等人将增强的 D* lite 算法应用于未知动态环境中的机器人路径导航领域 [8]。此外，这些方法通常与其他算法结合解决冲突问题，例如聚类方法 [9] 和勒让德伪谱方法 [10]。

(2) Optimal control theory methods

(2) 最优控制理论方法

The optimal control theory methods select the permissible control rate according to the kinematic model or the time domain mathematical model so that the UAV operates according to the constraints and thus achieves collision avoidance. These methods mainly include three aspects of mixed-integer linear programming, nonlinear optimization, and dynamic programming. Radmanesh et al. proposed fast-dynamic Mixed Integer Linear Programming (MILP) for the path planning of UAVs in various flight formations, focusing on avoiding typical UAVs from colliding with any intruder aircraft [11]. De Waen et al. targeted complex scenarios with multiple obstacles, and divided the MILP problem into many smaller MILP subproblems for trajectory modeling, which ensures the scalability of MILP to solve conflict problems [12]. Alonso-Ayuso et al. developed an exact mixed-integer nonlinear optimization model based on geometric construction for tackling the aircraft conflict detection and resolution problem [13].

最优控制理论方法根据运动学模型或时域数学模型选择允许的控制速率，使得无人机在约束条件下运行，从而实现避障。这些方法主要包括混合整数线性规划、非线性优化和动态规划三个方面。Radmanesh 等人提出了一种快速动态混合整数线性规划 (MILP) 方法，用于无人机在各种飞行编队中的路径规划，重点是避免无人机与任何入侵飞机发生碰撞 [11]。De Waen 等人针对具有多个障碍物的复杂场景，将 MILP 问题划分为许多较小的 MILP 子问题进行轨迹建模，确保了 MILP 解决冲突问题的可扩展性 [12]。

Alonso-Ayuso 等人基于几何构建开发了一种精确的混合整数非线性优化模型，用于解决飞机冲突检测与解决难题 [13]。

Heuristic-based search methods are reliable and effective for achieving collision avoidance and are able to resolve conflicts among small numbers of UAVs, which is the most common case in practice today. However, this method is not very suitable for mUAV conflicts, especially when the airspace is crowded, in which case the collision avoidance paths generated may suffer from secondary conflict problems. The optimal control methods take the minimum interval between UAVs as the optimization condition, and its relatively complex theory will lead to a decrease in anti-interference capability and an increase in computation, so it is not able to meet the real-time requirements.

基于启发式的搜索方法在实现避障方面可靠且有效，能够解决少量无人机之间的冲突，这是当今实践中最常见的情况。然而，这种方法不太适合多无人机 (mUAV) 冲突，尤其是在空域拥挤的情况下，此时生成的避障路径可能会遭受二次冲突问题。最优控制方法将无人机之间的最小间隔作为优化条件，其相对复杂的理论将导致抗干扰能力下降和计算量增加，因此无法满足实时性要求。

(3) Artificial intelligence methods

(3) 人工智能方法

The widespread use of artificial intelligence (AI) in recent years has provided new ideas and implementation paths. Reinforcement learning (RL) is the study of how an agent can interact with the environment to learn a policy that maximizes the expected cumulative reward for a task. When RL is combined with the powerful understanding ability of deep learning, it is clear that it possesses better decision-making efficiency than humans in a nearly infinite state space [14, 15] . The application of DPL to the field of UAV collision avoidance can solve the problems presented by the methods described above, while achieving better avoidance in urban airspace with variable environmental states and meeting strict real-time requirements.

近年来人工智能 (AI) 的广泛应用提供了新的思路和实施路径。强化学习 (RL) 是研究一个智能体如何与环境互动以学习一种策略，这种策略能够最大化任务预期的累积回报。当 RL 与深度学习的强大理解能力相结合时，显然在近乎无限的状态空间中，它比人类具有更高的决策效率 [14, 15] 。将深度强化学习 (DPL) 应用于无人机避障领域，可以解决上述方法中呈现的问题，同时实现在环境状态多变的城区空域中的更好避障，并满足严格的实时性要求。

The authors of [16, 17] developed a Q-learning algorithm to design the dynamic movement of UAVs, but there is no assurance that it can handle high-dimensional input data. Singla et al. proposed a deep recurrent Q-network with temporal attention to realize the indoor autonomous flight of a UAV [18]. In [19], the DDQN algorithm was applied to ship navigation to achieve multi-ship collision avoidance in crowded waters. Li et al. designed a tactical conflict resolution method for air logistics transportation based on the D3QN algorithm, enabling UAVs to successfully avoid non-cooperative targets [20]. The value-based algorithms (e.g., D3QN and DDQN) can adapt to complex state spaces, but cannot provide satisfactory solutions for continuous control problems. The emergence of police-based RL has solved such problems; one of the most widely used and mature approaches in practice is the DDPG algorithm proposed by DeepMind [21]. For example, references [22-25] addressed the trajectory optimization in a two UAV scenario based on the DDPG algorithm. Ribeiro et al. [22] utilized a geometric approach to model conflict detection between two UAVs and trained the agent in conjunction with the DDPG algorithm to generate a resolution strategy. The authors of [23,24] utilized the DDPG algorithm to solve the UAV path following problem, taking into account the conflict risks during movement. In [25], a proper heading angle was obtained using the DDPG algorithm before the aircraft reached the boundary of the sector to avoid collisions. Alternatively, Proximal Policy Optimization (PPO) methods can be used in aircraft collision avoidance, and have shown a certain level of performance [26].

[16, 17] 的作者开发了一种 Q 学习算法来设计无人机的动态移动，但无法保证其能够处理高维输入数据。Singla 等人提出了一种带有时间注意力的深度循环 Q 网络，以实现无人机在室内的自主飞行 [18]。在 [19] 中，将 DDQN 算法应用于船舶导航，以实现在拥挤水域中的多船避碰。李等人基于 D3QN 算法为航空物流运输设计了一种战术冲突解决方法，使无人机能够成功避开非合作目标 [20]。基于价值的算法 (例如，D3QN 和 DDQN) 能够适应复杂的状态空间，但不能为连续控制问题提供满意的解决方案。基于警察的强化学习方法的涌现解决了这些问题；在实践中最广泛使用且最成熟的方法之一是 DeepMind 提出的 DDPG 算法 [21]。例如，参考文献 [22-25] 基于 DDPG 算法解决了两个无人机场景中的轨迹优化问题。Ribeiro 等人 [22] 利用几何方法来建模两个无人机之间的冲突检测，并训练了与 DDPG 算法结合的智能体以生成解决策略。文献 [23,24] 的作者利用 DDPG 算法解决了无人机路径跟踪问题，同时考虑了

移动过程中的冲突风险。在 [25] 中，使用 DDPG 算法在飞机到达扇形区域边界之前获得适当的方向角以避免碰撞。另外，也可以在飞机避碰中使用近端策略优化 (PPO) 方法，并已显示出一定水平的性能 [26]。

In summary, although a variety of UAV collision avoidance methods have been developed based on DRL, there are still several gaps in terms of actual application: (1) Scholars have typically rasterized the entire airspace when designing the state space [16,17,22], which has some specific limitations: UAVs can only move to an adjacent raster, which limits the action dimensions of UAVs, and is not able to adequately reflect their flexibility. The use of a discrete state space would cause a waste of airspace resources, and the whole raster area may become a no-fly zone due to some small buildings, thus reducing the space available for UAV flights. (2) The dimensional advantage is also an important factor in measuring the performance of the method, with most existing UAV collision avoidance methods borrowing from ground traffic, and thus the dimensional range is limited to 2D, which does not match the actual operation situation in the airspace, while also limiting the UAV avoidance actions that can be selected [18,20,26]. (3) When the DRL theory is applied to mUAV collision avoidance, in the existing literature, only one UAV is regarded as the agent, and the other UAVs are regarded as dynamic obstacles without resolution ability [20]; their tracks are previously planned. In actual operation, conflicts may arise from any UAV, so each UAV should have the ability to make their own decisions.

总结来说，尽管基于深度强化学习 (DRL) 的无人机避障方法已经发展多样，但在实际应用中仍存在几个不足之处:(1) 学者们在设计状态空间时通常将整个空域栅格化 [16,17,22]，这种方法具有一定的局限性: 无人机只能移动到相邻的栅格，这限制了无人机的动作维度，并且不能充分反映它们的灵活性。使用离散状态空间会导致空域资源的浪费，而且由于一些小建筑的存在，整个栅格区域可能变成禁飞区，从而减少了无人机飞行的可用空间。(2) 维度优势也是衡量方法性能的重要因素，现有的多数无人机避障方法借鉴了地面交通，因此维度范围仅限于二维，这与空域中的实际操作情况不符，同时也限制了无人机可以选择的避障动作 [18,20,26]。(3) 当将 DRL 理论应用于多无人机 (mUAV) 避障时，在现有文献中，只有一个无人机被视为智能体，其他无人机被视为无解析能力的动态障碍物 [20]；它们的轨迹是预先规划的。在实际操作中，任何无人机都可能引发冲突，因此每个无人机都应该具备自主决策的能力。

1.2.Our Contributions

1.2. 我们的贡献

To solve the above problems, in this paper, a more practical collision avoidance method for mUAVs is developed. The primary contributions of this study are the following:

为了解决上述问题，本文开发了一种更适合多无人机 (mUAV) 的避障方法。本研究的主要贡献如下:

(1) In this paper, an adaptive decision-making framework for mUAV collision avoidance is proposed. The adaptive framework enables UAVs to autonomously determine the avoidance action to be taken in 3D space, providing the UAVs with more options for extrication strategies when faced with static or dynamic obstacles. This framework combines the conflict resolution pool in order to transform mUAV conflicts into UAV pairs for avoidance, controlling the computational complexity at the polynomial level, thereby providing a new idea for mUAV collision avoidance.

(1) 在本文中，提出了一个用于微型无人机 (mUAV) 避障的自适应决策框架。该自适应框架使无人机能够自主决定在三维空间中采取的避障动作，为无人机面对静态或动态障碍时提供了更多的解脱策略选项。该框架结合了冲突解决池，以将 mUAV 冲突转化为无人机对进行避障，控制计算复杂度在多项式级别，从而为 mUAV 避障提供了新的思路。

(2) A DRL model for UAV pairs is designed based on a continuous action space and state space, reflecting the maneuverability of UAVs and avoiding wastage of urban airspace resources. The model endows each UAV with decision-making ability, and utilizes a fixed sector format to explore conflicting obstacles, thus simplifying the state of the agent and making it more adaptable to dense urban building environments.

(2) 基于连续动作空间和状态空间，为无人机对设计了一种深度强化学习 (DRL) 模型，反映了无人机的机动性并避免了城市空域资源的浪费。该模型赋予每个无人机决策能力，并使用固定扇区格式探索冲突障碍，从而简化了代理的状态，使其更能适应密集的城市建筑环境。

(3) The DDPG algorithm is introduced to train the agent, and its convergence speed is enhanced by proposing the mechanism of destination area dynamic adjustment.

(3) 引入了 DDPG 算法来训练代理，并通过提出目的地区域动态调整机制来提高其收敛速度。

A summary of surveys related to DRL methods in the field of UAV collision avoidance is provided in Table 1. The table shows that this research represents the first study to resolve mUAV conflicts in a 3D environment based on continuous state and action space.

表 1 提供了关于无人机避障领域深度强化学习方法的相关调查总结。表格显示，这项研究是基于连续状态和动作空间在三维环境中解决 mUAV 冲突的第一个研究。

Table 1. Related work on conflict resolution based on DRL.

表 1. 基于 DRL 的冲突解决相关研究。

Ref.	Methods	Type	State and Action Space	Action	Multi-UAV	Environment
[16]	Q-leaning	Value-based	Discrete/Discrete	choosing from 4 possible directions	NO	2D
[17]	Q-leaning	Value-based	Discrete/Discrete	choosing from 7 possible directions	YES	3D
[18]	DDQN	Value-based	Continuous/Discrete	choosing from 3 possible directions	NO	2D
[20]	D3QN	Value-based	Continuous/Discrete	choosing acceleration; choosing yaw angular velocity choosing heading	YES	2D
[22]	DDPG	police-based	Discrete/Continuous	angle(max:15°/s); choosing acceleration (1.0 kts/s)	NO	2D
[23]	DDPG	police-based	Continuous/Continuous	choosing altitude, heading angle	NO	3D
[24]	DDPG	police-based	Continuous/Continuous	choosing heading angle $(-30°, 30°)$	NO	2D
[26]	PPO	police-based	Continuous/Continuous	choosing heading angle $(-30°, 30°)$, and speed $[0\ \text{m/s}, 40\ \text{m/s}]$	NO	2D
This paper	Improved DDPG	police-based	Continuous/Continuous	choosing velocity, heading angle, and altitude	YES	3D

参考文献	方法	类型	状态与动作空间	动作	多无人机 (UAV)	环境
[16]	Q 学习	基于价值的	离散/离散	从 4 个可能的方向中选择	否	2D
[17]	Q 学习	基于价值的	离散/离散	从 7 个可能的方向中选择	是	3D
[18]	双重深度 Q 网络 (DDQN)	基于价值的	连续/离散	从 3 个可能的方向中选择	否	2D
[20]	三维深度 Q 网络 (D3QN)	基于价值的	连续/离散	选择加速度；选择偏航角速度；选择航向	是	2D
[22]	深度确定性策略梯度 (DDPG)	基于警察的	离散/连续	角度 (最大:15°/秒)；选择加速度 (1.0 节/秒)	否	2D
[23]	深度确定性策略梯度 (DDPG)	基于警察的	连续/连续	选择高度，航向角	否	3D
[24]	深度确定性策略梯度 (DDPG)	基于警察的	连续/连续	选择航向角 $(-30°, 30°)$	否	2D
[26]	PPO(近端策略优化)	基于警察的	连续/连续	选择航向角 $(-30°, 30°)$ ，和速度 $[0\ \text{m/s}, 40\ \text{m/s}]$	否	2D
本文	改进的 DDPG(深度确定性策略梯度)	基于警察的	连续/连续	选择速度，航向角和高度	是	3D

The rest of the paper is organized as follows. In Section 2, the two-layer resolution framework is presented, and the methods for collision avoidance of UAV pairs and mUAVs are proposed. In Section 3, the improved DDPG algorithm is proposed for training the agent. In Section 4, the validity of the method is verified using a designed city scenario. In Section 5, some conclusions and summaries are presented. The table in the abbreviations section shows the main symbols used in this paper.

文章其余部分的组织如下。在第 2 节中，介绍了双层解决框架，并提出了无人机对和 mUAV 避障的方法。在第 3 节中，提出了改进的 DDPG 算法来训练代理。在第 4 节中，通过设计的城市场景验证了方法的有效性。在第 5 节中，提出了一些结论和总结。缩略词部分的表格显示了本文中使用的主要符号。

2. Problem Formulation

2. 问题公式化

In this section, we propose a two-layer resolution framework (Figure 1) and divide the problem formulation of mUAV collision avoidance into two parts: collision avoidance between UAV pairs, and collaborative mUAV collision avoidance. Firstly, the collision avoidance agent training model is designed based on DPL, and the agent can assign avoidance actions for UAV pairs in real time and complete collision avoidance for both dynamic obstacles (between UAVs) and static obstacles (buildings). Secondly, the collaborative mUAV collision avoidance model is proposed, in which mUAV conflicts are transformed into the form of UAV pairs and then the agent is used to deconflict them one by one.

在本节中，我们提出了一个双层解决框架 (图 1)，并将微型无人机 (mUAV) 避障问题公式分为两部分: 无人机对之间的避障和协同微型无人机避障。首先，基于 DPL 设计了避障代理训练模型，该代理能够为无人机对实时分配避障动作，并完成对动态障碍 (无人机之间) 和静态障碍 (建筑物) 的避障。其次，提出了协同微型无人机避障模型，其中将微型无人机冲突转换为无人机对的形式，然后使用代理逐一进行冲突解除。

The collision zero model of the UAV and building are generated as described in an existing study [27], and are elliptical and cylindrical shapes, as shown in Equations (1) and (2).

无人机的碰撞零模型与建筑物的生成如现有研究 [27] 所述，分别是椭圆形和圆柱形，如公式 (1) 和 (2) 所示。

$$D_u = \left\{ r \in R^3 : r^T A r \leq 1, A^{-1} = \operatorname{diag}\left({a_u}^2, {b_u}^2, {h_u}^2 \right) \right\} \tag{1}$$

$$D_o = \left\{ (x, y, z) \in R^3 : x^2 + y^2 \leq {R_o}^2, z \leq h_o \right\} \tag{2}$$

where a_u, b_u, h_u are the semi-axes of the ellipsoid, respectively, and R_o, h_o are the radius and height of the cylinder.

其中 a_u, b_u, h_u 是椭球体的半轴，分别地，R_o, h_o 是圆柱的半径和高度。

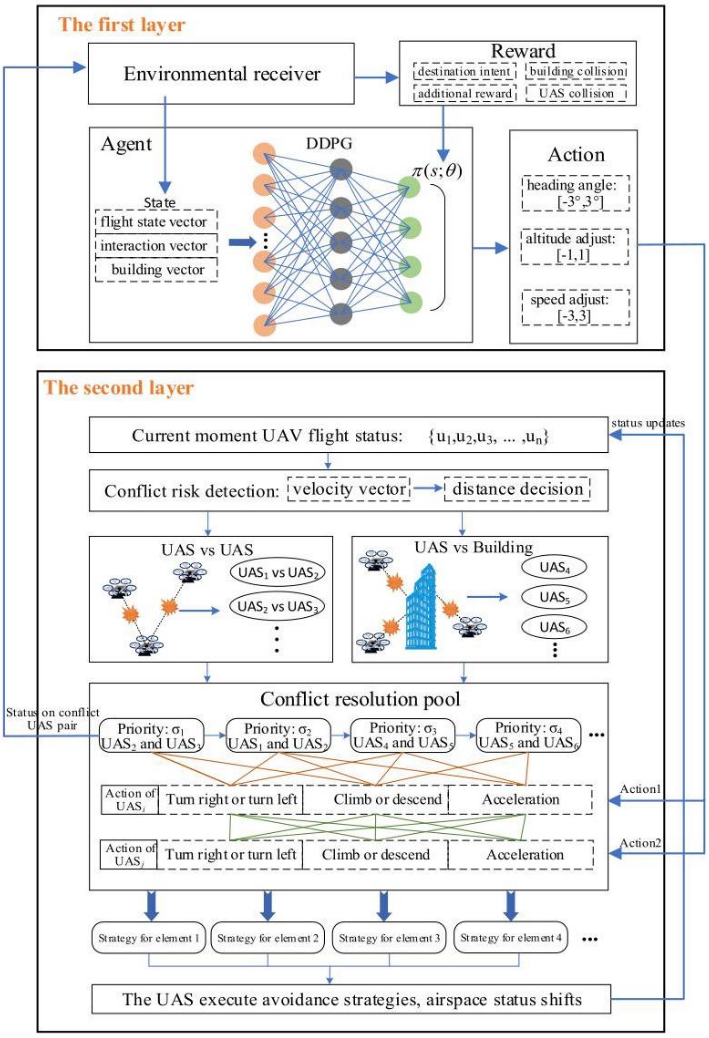


Figure 1. Overall workflow of the two-layer resolution framework.
图 1. 双层解决框架的整体工作流程。

2.1. First Layer: DRL-Based Method for Collision Avoidance between UAVs in a UAV Pair

2.1. 第一层: 基于 DRL 的无人机对之间避障方法

In this section, the first layer of the framework is designed, and the agent training model is constructed with respect to three aspects: state space, action space, and reward function.

在本节中，设计了框架的第一层，并根据状态空间、动作空间和奖励函数三个方面构建了代理训练模型。

2.1.1. Continuous State Space

2.1.1. 连续状态空间

According to the explanation in the previous section, we replace the rasterized space with a continuous state space. The agent state vector includes the following three parts:

根据上一节的解释，我们用连续状态空间代替栅格化空间。代理状态向量包括以下三个部分:

(1) The flight state vector of the UAV, including six attributes (φ : heading angle; V : horizontal speed; Z : the altitude of UAV; φ_g : relative heading angle of the destination to UAV; d_g : the horizontal distance of the destination to UAV), as shown in Figure 2. These attributes are able to accurately reflect the current flight status of the UAV, and the agent guides the UAV to its destination based on the flight status vector.

(1) 无人机的飞行状态向量，包括六个属性 (φ : 航向角；V : 水平速度；Z : 无人机的飞行高度；φ_g : 目的地相对于无人机的航向角；d_g : 目的地相对于无人机的水平距离)，如图 2 所示。这些属性能够准确反映无人机的当前飞行状态，并且代理基于飞行状态向量引导无人机到达目的地。

(2) Interaction vectors between UAV pairs (φ_{us} : the difference in heading angle; Z_{us} : the difference in altitude; d_{us} : the horizontal distance between two UAVs). These attributes reflect the position and heading relationships between the UAVs in a UAV pair, and the agent avoids collision between UAVs in a UAV pair based on the interaction vectors.

(2) 无人机对之间的交互向量 (φ_{us} : 航向角之差；Z_{us} : 高度之差；d_{us} : 两无人机之间的水平距离)。这些属性反映了无人机对中的无人机之间的位置和航向关系，并且代理基于交互向量避免无人机对之间的碰撞。

(3) Building vectors. There are lots of buildings in the urban airspace, and if all building information is put into the agent, this will result in high state vector dimensionality and affect the speed of convergence. In this paper, considering the detection range of the UAV in the horizontal direction, a flight sector is used to map obstacles affecting flight into a fixed-length vector, and these are regarded as the obstacle vectors, as shown in Figure 2.

(3) 建筑向量。在城市空域中有很多建筑物，如果将所有建筑信息都输入到代理中，这将导致状态向量维度过高，并影响收敛速度。在本文中，考虑到无人机在水平方向上的探测范围，使用飞行扇区将影响飞行的障碍物映射为固定长度的向量，这些被视为障碍物向量，如图 2 所示。

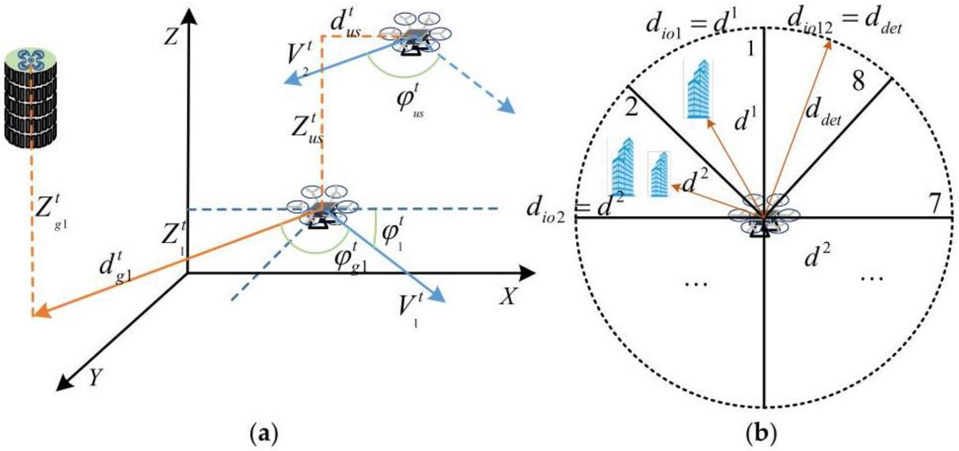


Figure 2. The diagram of the agent state vector. (a) Flight state vector. (b) Obstacle vectors.

图 2。代理状态向量图。(a) 飞行状态向量。(b) 障碍物向量。

The detection area is supposed to be a circle with the location as the center and the detection distance as the radius, that is, the agent can obtain the position information of obstacles in this area. Starting from the heading angle and rotating counterclockwise, every 45° is divided into one sector, and a total of eight sectors can be obtained. The distance between the UAV and all obstacles in each sector is calculated, and the closest distance is one of the attributes; if there are no obstacles in the sector, the value of this attribute corresponds to the detection distance. Supposing that there are j obstacles in sector m , then:

检测区域应为一个以位置为中心、检测距离为半径的圆，也就是说，代理能够获取该区域内障碍物的位置信息。从航向角开始逆时针旋转，每个 45° 被划分为一个扇区，总共可以获得八个扇区。计算无人机与每个扇区内所有障碍物的距离，最近距离是其中一个属性；如果扇区内没有障碍物，该属性的值对应于检测距离。假设扇区 j 中有 m 个障碍物，那么:

$$d^j_{\text{iom}} = \min\left(\left\|\left(x^j_{\text{iom}}, y^j_{\text{iom}}, z^j_{\text{iom}}\right) - (x^t_i, y^t_i, z^t_i)\right\| - R_u - R^j_o, d_{\text{det}}\right) \tag{3}$$

$$d_{\text{iom}} = \min\left(d^1_{\text{iom}}, d^2_{\text{iom}}, \cdots, d^j_{\text{iom}}\right)$$

where R_u denotes the collision zone radius of the UAV, R^j_o denotes the collision zone radius of building j, d_{det} denotes the UAV detection distance, d^j_{iom} denotes the distance attribute between UAV and building $j, d_{\text{ iom }}$ denotes the distance attribute of sector m , so the obstacle vector of UAV is: $[d_{io1}, d_{io2}, \ldots, d_{io8}]$.

其中 R_u 表示无人机的碰撞区域半径，R_o^j 表示建筑物碰撞区域半径，$j, d_{\text{ det }}$ 表示无人机检测距离，$d^j{}_{\text{iom}}$ 表示无人机与建筑物之间的距离属性，$j, d_{\text{ iom }}$ 表示扇区的距离属性，因此无人机的障碍物向量是: $[d_{io1}, d_{io2}, \ldots, d_{io8}]$ 。

In summary, the state vector received by the agent from the environment at moment t is:

总结来说，代理在时刻 t 从环境中接收到的状态向量是:

$$S_t = \left[\underbrace{\varphi_{1'}^t, V_{1'}^t, Z_1', \varphi_{g1}^t, Z_{g1}^t, d_{g1}^t, d_{g1}^t}_{\text{State of UAS2 Interaction of UAS2 Interaction of UAS2 Interaction of UAS2}}, \underbrace{\varphi_{us}^t, Z_{us}^t, d_{us}^t}_{\text{UAS}}, \underbrace{d_{101}^t, \ldots, d_{108}^t, d_{201}^t, \ldots, d_{208}^t}_{\text{Obstacles of UAS1 and UAS2}} \right], S_t \in S \tag{4}$$

2.1.2. Continuous Action Space

2.1.2. 连续动作空间

When modeling the UAV action space based on deep reinforcement learning, the control of the UAV is usually achieved based on adjusting spatial position, velocity, or acceleration. For the continuous state space and the flexibility of UAVs, in this paper, the continuous action space is designed based on adjusting the velocity and the maneuvering methods of UAVs are simplified, and are summarized into three processes: heading adjustment, altitude adjustment, and speed adjustment ($\Delta\varphi$: alteration in heading angle; ΔZ : alteration in altitude; ΔV : alteration in horizontal speed), where at each time step, $\Delta\varphi \in [-3°, 3°], \Delta Z \in [-1\text{ m}, 1\text{ m}], \Delta V \in [-2\text{ m/s}, 2\text{ m/s}]$. Thus, the action space of the agent at moment t is:

当基于深度强化学习建模无人机的动作空间时，通常通过调整空间位置、速度或加速度来实现对无人机的控制。由于连续状态空间和无人机的灵活性，本文设计了基于调整速度的连续动作空间，并将无人机的机动方法简化为三个过程: 航向调整、高度调整和速度调整 ($\Delta\varphi$: 航向角变化；ΔZ : 高度变化；ΔV : 水平速度变化)，在每个时间步长，$\Delta\varphi \in [-3°, 3°], \Delta Z \in [-1\text{ m}, 1\text{ m}], \Delta V \in [-2\text{ m/s}, 2\text{ m/s}]$ 。因此，代理在时刻 t 的动作空间是:

$$A_t = \left[\underbrace{\Delta\varphi_1, \Delta Z_1, \Delta V_1}_{\text{Action of UAS1}}, \underbrace{\Delta\varphi_2, \Delta Z_2, \Delta V_2}_{\text{Action of UAS2}} \right], A_t \in A \tag{5}$$

2.1.3. Reward Function Design

2.1.3. 奖励函数设计

The reward is a scalar feedback signal given by the environment that shows how well an agent performs at performing a certain strategy at a certain step. The reward function is a key component of the DRL framework. The purpose of the agent interacting with the environment is to maximize its reward value, so designing a suitable reward function for the agent can improve training performance and lead to faster convergence.

奖励是环境给出的一个标量反馈信号，它显示了代理在特定步骤执行特定策略的表现。奖励函数是深度强化学习框架的关键组成部分。代理与环境的互动目的是最大化其奖励值，因此为代理设计一个合适的奖励函数可以提高训练性能并导致更快的收敛。

In this paper, we consider the shortest path to the destination, avoid collision between UAV pair, avoid collision between UAV and building, avoid the UAV flying out of the specific area, set four reward functions, and use artificially designed "dense" rewards to achieve a dynamic balance between near-term and long-term rewards, which can mitigate the sparse reward problem in DRL.

在本文中，我们考虑了到达目的地的最短路径，避免无人机对之间的碰撞，避免无人机与建筑物的碰撞，避免无人机飞出特定区域，设定了四个奖励函数，并使用人工设计的“密集”奖励来实现短期奖励与长期奖励之间的动态平衡，这可以缓解深度强化学习中的稀疏奖励问题。

(1) Destination intent reward: when there are no obstacles in the sector of the UAV pair, the destination intent reward is used to ensure that the UAV takes the shortest path to the destination. The

entire movement of the UAV is divided into multiple ”intensive” actions, and the reward function is set for them to ensure that each action of the UAV affects the final reward value, thus contributing to the improvement of the overall strategy, as shown in Equation (6):

(1) 目的地意图奖励: 当无人机对的扇区中没有障碍物时，使用目的地意图奖励来确保无人机沿最短路径飞向目的地。将无人机的整个移动过程分为多个“密集”动作，并为它们设置奖励函数以确保无人机的每个动作都会影响最终奖励值，从而有助于提高整体策略，如方程 (6) 所示:

$$r_1^i = \begin{cases} 0.5 & \varphi_{gi}^t \in [0°, 10°] \text{ or } \varphi_{gi}^t \in [350°, 360°] \\ 0.1 & \varphi_{gi}^t \in (10°, 20°) \text{ or } \varphi_{gi}^t \in [340°, 350°) \\ 0 & \varphi_{gi}^t \in (20°, 90°] \text{ or } \varphi_{gi}^t \in [270°, 340°) \\ -0.1 & \text{else} \end{cases} \tag{6}$$

$$r_2^i = \left(d_{gi}^{t-1} - d_{gi}^t\right) + \left(Z_{gi}^{t-1} - Z_{gi}^t\right)$$

$$R_d^i = r_1^i + r_2^i$$

(2) Building collision avoidance reward: when there are obstacles in the sector of the UAV pair, it is necessary to ensure that the UAVs avoid colliding with buildings while flying to their destinations; therefore, we need to balance the two tasks, and the reward function at this time is shown in Equation (7).

(2) 建筑物碰撞避免奖励: 当无人机对的扇区中存在障碍物时，需要确保无人机在飞向目的地的过程中避免与建筑物碰撞；因此，我们需要平衡这两个任务，此时的奖励函数如方程 (7) 所示。

$$r_1^i = \begin{cases} 0.5 & \varphi_{gi}^t \in [0°, 10°] \text{ or } \varphi_{gi}^t \in [350°, 360°] \\ 0.1 & \varphi_{gi}^t \in (10°, 20°] \text{ or } \varphi_{gi}^t \in [340°, 350°) \\ 0 & \text{else} \end{cases}$$

$$r_2^i = \left(d_{gi}^{t-1} - d_{gi}^t\right) + \left(Z_{gi}^{t-1} - Z_{gi}^t\right) \tag{7}$$

$$r_3^i = \left(d^t{}_{\text{iom}} - d_{\text{iom}}^{t-1}\right) + \left(d^t{}_{\text{iom}} / d_{\text{ det}} - 1\right)$$

$$R_d^i = r_1^i + r_2^i + r_3^i$$

(3) UAV collision avoidance reward: the UAV collision avoidance reward is used to avoid collisions between the UAV and other UAVs. A UAV alert area is set up, and UAV collision avoidance is made the main task when there are other UAVs within the alert area, as shown in Equation (8).

(3) 无人机碰撞避免奖励: 无人机碰撞避免奖励用于避免无人机与其他无人机之间的碰撞。设置了一个无人机警示区域，当警示区域内有其他无人机时，将无人机碰撞避免作为主要任务，如方程 (8) 所示。

$$r_1^{ij} = \begin{cases} -0.1 \cdot \left(\frac{\overrightarrow{V}_i^t \cdot \overrightarrow{V}_j^t}{\left|\overrightarrow{V}_i^t\right| \cdot \left|\overrightarrow{V}_j^t\right|}\right) - 1 & \underbrace{d_{us}^t < d_{det} \text{ and } Z_{us}^t < 4 \cdot H_u}_{\text{Alerting Zone}} \\ 0 & \text{else} \end{cases} \tag{8}$$

$$r_2^{ij} = \begin{cases} \left(d_{us}^t - d_{us}^{t-1}\right) + \left(Z_{us}^t - Z_{us}^{t-1}\right) + \left(\frac{d_{us}^t}{d_{det}} + \frac{Z_{us}^t}{4 \cdot H_u} - 2\right) & \text{Alerting} \\ 0 & \text{else} \end{cases}$$

$$R_{uu}^{ij} = r_1^{ij} + r_2^{ij}$$

where $\overrightarrow{V}_i^t$ is the direction vector of the velocity of the UAV at moment t .

其中 $\overrightarrow{V}_i^t$ 是无人机在时刻 t 的速度方向向量。

(4) Additional reward: There are four final states of the UAV: reaching the destination, flying out of the control area, colliding with obstacles, and colliding with other drones. This additional reward is used to provide a relatively large reward or penalty value when the UAV reaches its final state, which can be used to guide the UAV to its destination and avoid bad events such as collisions or loss of control, as shown in Equation (9).

(4) 额外奖励: 无人机的最终状态有四种: 到达目的地、飞出控制区域、与障碍物碰撞以及与其他无人机碰撞。这种额外奖励用于在无人机达到其最终状态时提供相对较大的奖励或惩罚值，这可以用来引导无人机到达目的地并避免碰撞或失控等不良事件，如方程 (9) 所示。

$$R_{ex}^{i}=\begin{cases}12: U_{des}^{i}\in D_{u}^{i}\\ -6: \text{ Drive out of control area}\\ -6: D_{o}^{j}\cap D_{u}^{i}\neq\varnothing\\ -6: D_{u}^{j}\cap D_{u}^{i}\neq\varnothing\end{cases} \tag{9}$$

In summary, the reward function received by the agent per unit of time is as follows:
总结来说，智能体每单位时间接收到的奖励函数如下:

$$R_t=R_d^i+R_{uu}^{ij}+R_{ex}^{i}\ \forall i,j\in\{1,2\} \tag{10}$$

2.1.4.The Interaction between the Agent and the Environment

2.1.4. 智能体与环境的交互

In reinforcement learning, the interaction between the agent and its environment is often modeled by a Markovian decision process. This process can be represented by a four-tuple (S,A,R,γ) , where S is the current state of the agent Equation (4), A is the action taken by the agent (Equation (5)), R is the reward value obtained by the agent after taking the current action, $\gamma\in[0,1]$ is a discount factor, which is a constant, as shown in Figure 3.

在强化学习中，智能体与其环境的交互通常通过一个马尔可夫决策过程来建模。这个过程可以用一个四元组 (S,A,R,γ) 表示，其中 S 是智能体的当前状态 (方程 (4))，A 是智能体采取的动作 (方程 (5))，R 是智能体在采取当前动作后获得的奖励值，$\gamma\in[0,1]$ 是折扣因子，这是一个常数，如图 3 所示。

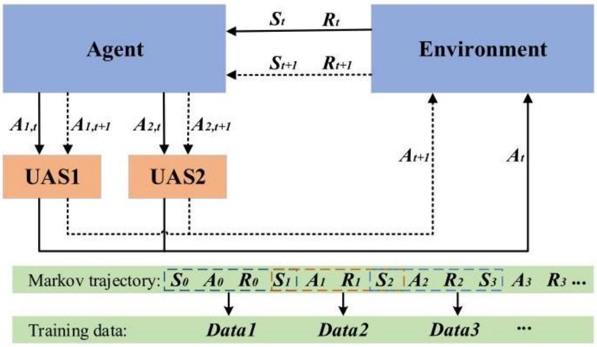

Agent
Environment
S_t R_t
S_{t+1} R_{t+1}
$A_{1,t}$
$A_{1,t+1}$
$A_{2,t}$
$A_{2,t+1}$
UAS1
UAS2
A_{t+1}
A_t
Markov trajectory: S_0 A_0 R_0 S_1 A_1 R_1 S_2 A_2 R_2 S_3 A_3 R_3 ...
Training data: Data1 Data2 Data3 ...

Figure 3. The interaction mode between the agent and the environment.
图 3. 智能体与环境的交互模式。

Assuming a discrete time domain $t\in\{0,1,2,\cdots\}$, the agent starts from the initial state S_0 and observes the environment state $S_t\in S$ at a certain time node t , taking action $A_t\in A$ based on the state and specific decision-making policy, and the next state is transferred to S_{t+1} , while the agent receives an immediate reward $R_t=R\left(S_t,A_t\right)$. All the variables obtained from the Markov decision process can be recorded as a trajectory $\tau=\left(S_0,A_0,R_0,S_1,A_1,R_1,\cdots\right)$, and each segment of the trajectory can be intercepted to form a set of training data for the subsequent training of the Target-network and Evaluate-network in the algorithm.

假设一个离散时间域 $t\in\{0,1,2,\cdots\}$ ，智能体从初始状态 S_0 开始，在特定时间节点 t 观察环境状态 $S_t\in S$ ，根据状态和特定的决策策略采取动作 $A_t\in A$ ，下一个状态转移到 S_{t+1} ，同时智能体接收到即时奖励 $R_t=R\left(S_t,A_t\right)$ 。所有从马尔可夫决策过程中获得的变量可以记录为一个轨迹 $\tau=\left(S_0,A_0,R_0,S_1,A_1,R_1,\cdots\right)$ ，轨迹的每个片段可以被截取形成一组训练数据，用于算法中目标网络和评估网络的后续训练。

2.2. Second Layer: Collaborative Collision Avoidance for mUAVs

2.2. 第二层: 多无人机协同避障

2.2.1. Three-Dimensional Conflict Detection

2.2.1. 三维冲突检测

To ensure mUAV safety, the conflict risk of each UAV needs to be detected, and then targeted for resolution. The conflict risk is detected based on the velocity vector and distance; the risk includes two types: UAV vs. UAV and UAV vs. building.

为了确保微型无人机的安全，需要检测每架无人机的冲突风险，并进行针对性的解决。冲突风险的检测基于速度向量和距离；风险包括两种类型: 无人机与无人机之间的冲突以及无人机与建筑物之间的冲突。

For conflict detection between the UAVs in a UAV pair, subscripts S and R designate the stochastic UAV and the reference UAV, respectively, in any UAV pair, P_R denotes the position of the reference UAV, P_S denotes the position of the stochastic UAV, and the velocities of the reference UAV and the stochastic UAV at the current moment are V_R, $\mathrm{V_S}$.

对于无人机对之间的冲突检测，下标 S 和 R 分别表示随机无人机和参考无人机，在任何无人机对中，P_R 表示参考无人机的位置，P_S 表示随机无人机的位置，当前时刻参考无人机和随机无人机的速度分别为 V_R, $\mathrm{V_S}$ 。

In the process of modeling, defining a combined collision zone, which is assigned to the reference UAV so that the stochastic UAV can be regarded as a particle. The parameters of combined collision region D are:

在建模过程中，定义一个组合碰撞区域，并将其分配给参考无人机，以便将随机无人机视为一个粒子。组合碰撞区域 D 的参数为:

$$H_D = 2H_u = 2\sqrt{3}h_u \; R_D = 2R_u = 2\sqrt{3}a_u \tag{11}$$

A 3D collision coordinate system is established with the origin fixed at the position of the reference UAV; the relative position and velocity of the UAV pair are: $\Delta P_{uu} = P_S - P_R$, $\Delta V_{uu} = V_S - V_R$. It is assumed that Δl is the extension of the relative velocity ΔV_{uu} , and if the intersection of Δl and the combined collision region D is non-empty, i.e., $\Delta l \cap D \neq \varnothing$, then the UAV pair meets the condition of conflict in the relative velocity dimension; then, the spatial distance of the UAV pair is calculated, and if the spatial distance is less than the threshold while satisfying the relative velocity conflict condition, it can be determined that there is a collision risk for this UAV pair.

建立一个以参考无人机位置为原点的三维碰撞坐标系；无人机对的相对位置和速度为: $\Delta P_{uu} = P_S - P_R$ ，$\Delta V_{uu} = V_S - V_R$ 。假设 Δl 是相对速度 ΔV_{uu} 的延伸，如果 Δl 与组合碰撞区域 D 的交集非空，即 $\Delta l \cap D \neq \varnothing$ ，则该无人机对满足相对速度维度下的冲突条件；然后，计算无人机对的空间距离，如果空间距离小于阈值且满足相对速度冲突条件，则可以确定这对无人机存在碰撞风险。

For conflict detection between UAVs and buildings, the combined collision zone is also defined. Since buildings are fixed, it is only necessary to consider whether the velocity extension line of the UAV intersects the combined collision region, and then the spatial distance between the UAV and building is considered to determine the conflict, as shown in Figure 4.

用于检测无人机与建筑物之间冲突的联合碰撞区域也被定义。由于建筑物是固定的，因此只需考虑无人机的速度延长线是否与联合碰撞区域相交，然后考虑无人机与建筑物之间的空间距离来确定是否存在冲突，如图 4 所示。

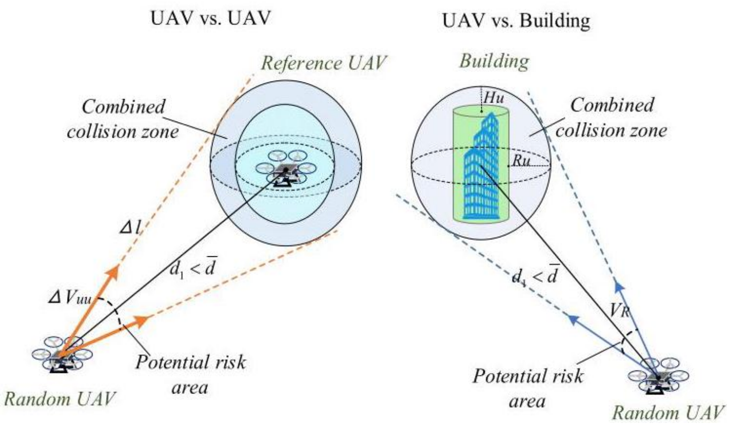


Figure 4. The diagram of three-dimensional conflict detection.
图 4. 三维冲突检测示意图。

2.2.2. Conflict Resolution Pool

2.2.2. 冲突解决池

Based on the UAV conflict detection method in the previous section, we propose the concept of a conflict resolution pool, where the UAVs detected as being at risk are stored as the objects of the collision avoidance agent. For UAV-to-UAV conflicts, the elements in the pool are UAV pairs, and for UAV-to-building conflicts, the elements are single UAVs, while the distances of UAVs at risk are deposited in the pool as the priorities of conflict resolution, with smaller distances indicating higher levels of risk and a more urgent need for collision avoidance, as shown in Equation (12).

基于上一节中提出的无人机冲突检测方法，我们提出了冲突解决池的概念，其中将检测为有风险的无人机存储为避障代理的对象。对于无人机之间的冲突，池中的元素是无人机对；对于无人机与建筑物的冲突，元素是单个无人机，而有风险的无人机的距离则作为冲突解决的优先级存储在池中，距离越小表示风险越高，需要更紧急地进行避障，如公式 (12) 所示。

$$P = \left\{ \underbrace{(u_1, u_2) : \sigma_{12}, \cdots ,}_{\text{Setelements}}, \underbrace{(u_i, u_j) : \sigma_{ij}, a_3 : \sigma_3, \cdots , u_l : \sigma_l}_{\text{Conflict drone pairs}} \right\} \sigma_{ij} = \frac{1}{d_{ij}}, \sigma_l = \frac{1}{d_l} \tag{12}$$

where (u_i, u_j) is the conflict UAV pair, u_l is the UAV in conflict with the building, σ is the priority, and d is the spatial distance.

其中 (u_i, u_j) 是冲突的无人机对，u_l 是与建筑物发生冲突的无人机，σ 是优先级，d 是空间距离。

The conflict resolution pool transforms mUAV conflicts into UAV pairs for avoidance, which simplifies the cooperative collision avoidance problem to a great extent. If a trajectory search-based approach is used to calculate the safe path for each UAV individually, the search space will grow exponentially with the number of UAVs, and when the number exceeds a certain value, the complexity of the algorithm will be too high to be able to solve the problem within an acceptable time. The conflict resolution pool prevents the resolution energy being wasted on temporarily safe UAVs, so that the computational complexity of the method can be controlled at the polynomial level, and the ability to perform collision avoidance will be further improved.

冲突解决池将多无人机冲突转化为无人机对进行避障，这在很大程度上简化了协同避障问题。如果使用基于轨迹搜索的方法为每个无人机单独计算安全路径，搜索空间将随着无人机数量的增加而指数增长，当数量超过一定值时，算法的复杂度将过高，无法在可接受的时间内解决问题。冲突解决池避免了将解决能量浪费在暂时安全的无人机上，因此可以将方法的计算复杂度控制在多项式级别，并进一步提高避障能力。

2.2.3. Collaborative Resolution Process for mUAVs

2.2.3. 多无人机协同解决过程

In this section, we propose a working model for this method by combining the concepts of the collision avoidance agent, three-dimensional conflict detection, and the conflict resolution pool.

在本节中，我们通过结合避障代理、三维冲突检测和冲突解决池的概念，提出了这种方法的工作模型。

Assuming that there are n UAVs in the urban airspace, the specific steps are as follows: Step 1: A pool K is built, consisting of all UAVs in the airspace, which is initialized for each timestamp:

假设在城市空域中有 n 架无人机，具体步骤如下: 步骤 1: 建立一个包含空域中所有无人机的池 K ，在每个时间戳对其进行初始化:

$$K = \{u_1, u_2, u_3, \cdots, u_n\} \tag{13}$$

Step 2: The three-dimensional conflict detection method is used to detect UAVs at risk, which are stored in the conflict resolution pool S , and then the priorities of the pool elements are calculated.

步骤 2: 使用三维冲突检测方法检测处于风险的无人机，将这些无人机存储在冲突解决池 S 中，然后计算池元素的优先级。

Step 3: The UAV pair with the highest priority is selected, as follows:

步骤 3: 选择优先级最高的无人机对，如下:

$$(i, j) = \underset{(i,j)}{\operatorname{argmax}} \left(\sigma_{ij}\right) \tag{14}$$

If the UAV pairs are all in the pool K , the reinforcement learning agent is used to assign avoidance actions to them. For UAVs that are not in K , the actions assigned to it by the agent are ignored, and the original actions are kepts unchanged. Meanwhile, this UAV pair is removed from pool S .

如果无人机对都在池 K 中，则使用强化学习代理为它们分配避障动作。对于不在 K 中的无人机，代理分配给它的动作将被忽略，保持原动作不变。同时，该无人机对从池 S 中移除。

Step 4: When there are no UAV pairs in the conflict resolution pool S , the two UAVs with the highest priorities that are in conflict with the building are selected and a UAV pair is formed, as follows:

步骤 4: 当冲突解决池 S 中没有无人机对时，选择与建筑物冲突的优先级最高的两架无人机，形成一个无人机对，如下:

$$(m, n) : m = \underset{l}{\operatorname{argmax}} \left(\sigma_l\right), n = \underset{l'}{\operatorname{argmax}} \left(\sigma_{l'}\right), m \neq n \tag{15}$$

The agent is used to assign avoidance actions to them, and the corresponding UAVs are removed from conflict resolution pool S .

使用代理为它们分配避障动作，并将相应的无人机从冲突解决池 S 中移除。

Step 5: The UAVs that have been assigned avoidance actions are removed from pool K until $S = \varnothing$, and for the UAVs that are still in pool K , their original actions are kept unchanged.

步骤 5: 已分配避障动作的无人机从池 K 中移除，直到 $S = \varnothing$ ，对于仍停留在池 K 中的无人机，它们的原动作保持不变。

Figure 5 shows the process of collaborative collision avoidance.

图 5 展示了协同避障的过程。

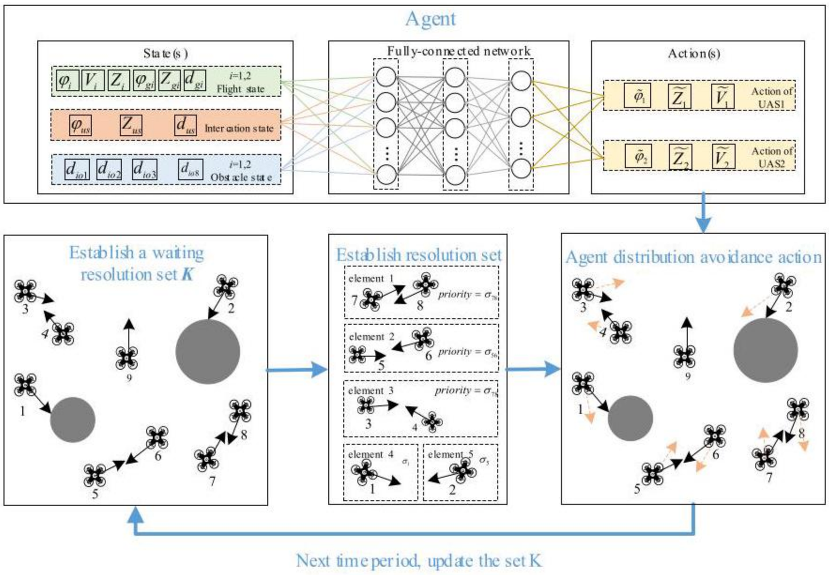

Agent
State(s)
φ_i V_i Z_i φ_{gi} Z_{gi} d_{gi}
i=1,2
Flight state
φ_{us} Z_{us} d_{us}
Interaction state
d_{io1} d_{io2} d_{io3} d_{io8}
i=1,2
Obstacle state
Fully-connected network
Action(s)
$\tilde{\varphi}_1$ $\widetilde{Z}_1$ $\widetilde{V}_1$
Action of UAS1
$\tilde{\varphi}_2$ $\widetilde{Z}_2$ $\widetilde{V}_2$
Action of UAS2
Establish a waiting resolution set K
Establish resolution set
element 1
priority = σ_{78}
element 2
priority = σ_{56}
element 3
priority = σ_{34}
element 4
σ_1
element 5
σ_2
Agent distribution avoidance action
1
2
3
4
5
6
7
8
9
Next time period, update the set K

Figure 5. Diagram of collaborative collision resolution.
图 5. 协同冲突解决示意图。

3. Improved Algorithm for Agent Training

3. 代理训练的改进算法

3.1. Deep Deterministic Policy Gradient

3.1. 深度确定性策略梯度

Faced with a continuous state space and action space, in this paper, the DDPG algorithm is taken to train the collision avoidance agent for the UAV pair. The DDPG algorithm utilizes four neural networks in the actor-critic framework: policy network $(\pi(s;\theta))$, Q network $(Q(s,a;\omega))$, target policy network $(\pi(s;\theta^{-}))$, and target Q network $(Q(s,a;\omega^{-}))$.

面对连续的状态空间和动作空间，本文采用 DDPG 算法来训练无人机对的避障代理。DDPG 算法在演员-评论家框架中使用了四个神经网络: 策略网络 $(\pi(s;\theta))$, Q 网络 $(Q(s,a;\omega))$ ，目标策略网络 $(\pi(s;\theta^{-}))$ ，和目标 Q 网络 $(Q(s,a;\omega^{-}))$ 。

The actor calculates the optimal action for the current state based on the learned policy function $(\pi_\theta(s_i))$. The critic estimates the value function $(Q_\omega(s,a))$ given the state and the action, which provides an expected accumulated future reward for this state-action pair. In addition, the critic is responsible for calculating the loss function (i.e., TD error) that is used in the learning process for both the policy network and the Q -network. To update the critic network, similar to Q -learning, the Bellman equation [28] is used:

演员根据学习到的策略函数 $(\pi_\theta(s_i))$ 计算当前状态的最优动作。评论家估计给定状态和动作的价值函数 $(Q_\omega(s,a))$ ，为这个状态-动作对提供预期的累积未来奖励。此外，评论家负责计算用于策略网络和 Q -网络学习过程中的损失函数 (即 TD 误差)。为了更新评论家网络，类似于 Q -学习，使用贝尔曼方程 [28] :

$$\text{target}_{t} = R_t + \gamma Q\left(S_{t+1}, \pi\left(S_{t+1}; \theta^{-}\right); \omega^{-}\right) \tag{16}$$

Then, the loss function is defined, and the argument is updated to minimize the loss between the original Q and the target:

然后，定义损失函数，并更新参数以最小化原始 Q 和目标之间的损失:

$$\text{Loss} = \frac{1}{n}\sum_{t=1}^{n}\left(\text{target}_t - Q\left(S_t, a_t; \omega\right)\right)^2 \tag{17}$$

The actor utilizes the policy network $(\pi(s;\theta))$ to select the best action, which maximizes the value function. The objective function in updating the actor is to maximize the expected return:

演员利用策略网络 $(\pi(s;\theta))$ 选择最佳动作，该动作最大化价值函数。更新演员的目标函数是最大化预期回报:

$$J(\theta) = \mathbb{E}\left(Q(s, a \mid \omega) \mid_{s=s_t, a=\pi_\theta(s_t)}\right) \tag{18}$$

According to the chain rule, the gradient of the objective function Q concerning the actor parameters can be obtained:

根据链式法则，可以得到关于演员参数的目标函数 Q 的梯度:

$$\nabla_\theta J(\theta) \approx \nabla_\theta \pi_\theta(s) \nabla_a Q(s, a) \tag{19}$$

Then, for mini-batch data, the mean of the sum of gradients is taken:

然后，对于小批量数据，取梯度和的平均值:

$$\nabla_\theta J(\theta) \approx \frac{1}{n} \sum_i \nabla_\theta \pi_\theta(s)|_{si} \nabla_a Q_\omega(s, a)\Big|_{s=si, a=\pi_\theta(si)} \tag{20}$$

The target network is a network used in the training phase. This network is equivalent to the original network being trained, and it provides the target values used to compute the loss function. In the DDPG algorithm, the target network is modified using a soft update:

目标网络是在训练阶段使用的网络。这个网络等同于正在训练的原始网络，它提供了用于计算损失函数的目标值。在 DDPG 算法中，目标网络通过软更新进行修改:

$$\begin{cases} \theta^- \leftarrow \tau\theta + (1-\tau)\theta^- \\ \omega^- \leftarrow \tau\omega + (1-\tau)\omega^- \end{cases} \tag{21}$$

This means that the target weights are constrained to change slowly. The use of target networks with soft updates allows them to give consistent targets during the temporal-difference backups and causes the learning process to remain stable

这意味着目标权重被限制为缓慢变化。使用具有软更新的目标网络可以在时间差备份期间提供一致的目标，并使学习过程保持稳定。

3.2.An Improved Measure for DDPG

3.2. DDPG 的改进度量

The DDPG algorithm has high execution efficiency, enabling continuous motion control of the agent. However, presented with the specific environment described in this paper, the DDPG consumes too much time in agent training, making it difficult to respond quickly when the urban environment undergoes significant changes and the agent needs to be retrained. To address such problems, this section improves the algorithm mainly in terms of the dynamic adjustment of the destination area.

DDPG 算法具有高执行效率，能够实现代理的连续运动控制。然而，在本文描述的特定环境中，DDPG 在代理训练上消耗过多时间，使得在城市环境发生重大变化且代理需要重新训练时，难以快速响应。为了解决此类问题，本节主要从目标区域的动态调整方面改进算法。

Due to the large spatial area of the city, the UAV destination is relatively small, and is replaced by a prime point in Equation (9), where $U_{des}^i \in D_u^i$ indicates that the UAV has reached its destination, and the agent receives the corresponding reward. However, in the actual training process, it is difficult to achieve the above conditions when the UAV performs the search, so there is little chance for the agent to obtain a relatively large reward value, thus causing the convergence speed to decrease.

由于城市空间范围较大，无人机目的地相对较小，并且在公式 (9) 中由一个质点代替，其中 $U_{des}^i \in D_u^i$ 表示无人机已到达目的地，代理接收到相应的奖励。然而，在实际训练过程中，当无人机执行搜索时，很难达到上述条件，因此代理获得较大奖励值的机会很小，从而导致收敛速度降低。

In this section, the way of obtaining destination rewards in the algorithm is improved based on the Wright learning curve model, and a dynamic adjustment mechanism for the destination area is proposed. During the early stage of training, the destination area is expanded so that the agent can complete the task relatively easily and learn the primary strategy. According to the learning curve, the destination area is gradually reduced, and the agent gradually learns more difficult strategies, which is conducive to improving the stability of the learning and accelerating the convergence speed of the algorithm. The destination range is defined as being spherical:

在本节中，基于 wright 学习曲线模型，改进了算法中获取目的地奖励的方式，并提出了目标区域的动态调整机制。在训练早期，扩大目标区域，使代理可以相对容易地完成任务并学习主要策略。根据学习曲线，逐渐缩小目标区域，代理逐渐学习更困难的策略，这有利于提高学习的稳定性和加速算法的收敛速度。目标范围被定义为球形:

$$U^{i}{}_{\text{des}} \in D^{i}{}_{\text{des}}\, , D^{i}{}_{\text{des}} = \left\{ r \in R^{3} : r^{T} r \leq R^{i}{}_{\text{des}} \right\} \tag{22}$$

In Equation (22), $D^{i}{}_{\text{des}}$ represents the destination area, if $D_{u}^{i} \cap D^{i}{}_{\text{des}} \neq \varnothing$ means the UAV has reached its destination, the area radius is adjusted with the training episode according to the Wright learning curve model:

在公式 (22) 中，$D^{i}{}_{\text{des}}$ 代表目标区域，如果 $D_{u}^{i} \cap D^{i}{}_{\text{des}} \neq \varnothing$ 表示无人机已到达目的地，则根据 wright 学习曲线模型，训练剧集调整区域半径:

$$\begin{cases} \alpha = \frac{\lg C}{\lg 2} \\ R^{i}{}_{\text{des}} = \bar{R} \cdot x^{\alpha} \end{cases} \tag{23}$$

In Equation (23), α is the learning rate, C is the attenuation coefficient, $\bar{R}$ is the initial destination area radius, and x indicates the training episode. The dynamic adjustment mechanism of the destination area further optimizes the "dense" reward and allows the algorithm to learn useful experiences in the early stages.

在公式 (23) 中，α 是学习率，C 是衰减系数，$\bar{R}$ 是初始目标区域半径，x 表示训练阶段。目标区域的动态调整机制进一步优化了“密集”奖励，并允许算法在早期阶段学习有用的经验。

The original DDPG algorithm needs more training epochs to detect the accurate location due to the small and fixed destination area, and sometimes even fails to obtain the destination reward. The improved DDPG algorithm adds a dynamic adjustment mechanism for the destination area, which enlarges the size of the destination area in the initial stage of training, so that the agent can easily obtain the approximate destination location, ensuring that it will move in the right direction in the subsequent training. As the training progresses, the algorithm gradually reduces the destination area based on the Wright learning curve, guiding the agent to the precise destination location. Compared with the original algorithm, the improved DDPG algorithm is more goal oriented and avoids ineffective exploration on the part of the agent, so it can accelerate the convergence speed and save training resources.

原始的 DDPG 算法由于目标区域小且固定，需要更多的训练周期来检测准确的位置，有时甚至无法获得目标奖励。改进的 DDPG 算法为目的地区域增加了一个动态调整机制，它在训练的初始阶段扩大了目标区域的大小，使得智能体可以轻松获得近似的目标位置，确保在后续训练中它会向正确的方向移动。随着训练的进行，算法根据 wright 学习曲线逐渐减小目标区域，引导智能体到达精确的目标位置。与原始算法相比，改进的 DDPG 算法更具目标导向性，避免了智能体进行无效探索，因此可以加速收敛速度并节省训练资源。

4. Results and Discussion

4. 结果与讨论

4.1. Environment Setting and Hyperparameters

4.1. 环境设置和超参数

To analyze the performance of the mUAV collision avoidance method, in this paper, a DJI Matrice 600 is selected as a case study, whose form factor (L × W × H) is set to 1668 mm × 1668 mm × 759 mm , and the calculated elliptical collision zero parameters are 1445 mm × 1445 mm × 657 mm . The scenario range is set to 1000 m × 1000 m × 50 m , considering that the collision avoidance area of small UAVs in the city will not be too large.

为了分析 mUAV 避障方法的性能，本文选择 DJI Matrice 600 作为案例研究，其形态因子 (L × W × H) 设置为 1668 mm × 1668 mm × 759 mm ，计算出的椭圆形避障零参数为 1445 mm × 1445 mm × 657 mm 。场景范围设置为 1000 m × 1000 m × 50 m ，考虑到城市中小型无人机的避障区域不会太大。

In the experimental scenario, we construct the spatial layout of buildings in the city and set up fixed-volume obstacles at fixed locations; the building collision zero is cylindrical, and the shape parameters are shown in Table 2. The experiment was based on eight UAVs, each with an initial speed and initial heading. At the beginning of training, a random origin is generated for each UAV and the terminal, and the agent assigns avoidance actions to UAVs according to the action space in Equation (5), and returns to the origin to restart the training if a UAV has a collision accident. The experimental parameters are as shown in Table 3, and the environment is as shown in Figure 6.

在实验场景中，我们构建了城市中建筑物的空间布局，并在固定位置设置了固定体积的障碍物；建筑碰撞零是圆柱形的，其形状参数如表 2 所示。实验基于八架无人机，每架无人机具有初始速度和初始航向。在训练开始时，为每个无人机和终端生成一个随机起点，并根据方程 (5) 中的动作空间为无人机分配避障动作，如果无人机发生碰撞事故，则返回起点重新开始训练。实验参数如表 3 所示，环境如图 6 所示。

Table 2. The shape parameters of the building.

表 2. 建筑的形状参数。

Obstacle Number	X (m)	Y (m)	R (m)	Z (m)	Obstacle Number	X (m)	Y (m)	R (m)	Z (m)
1	500	500	80	100	6	245	458	30	89
2	200	200	15	70	7	660	150	40	56
3	900	567	25	85	8	900	328	22	78
4	850	820	35	80	9	326	895	17	78
5	150	698	18	60	10	628	736	20	91

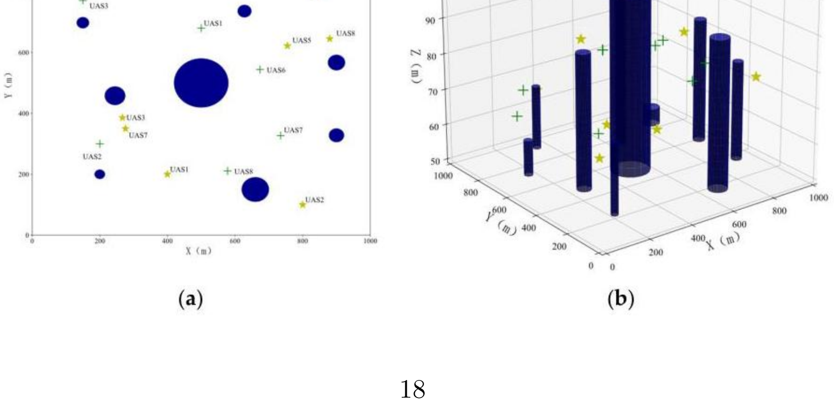
障碍物数量	X (m)	Y (米)	R (m)	Z (米)	障碍物数量	X (m)	Y (米)	R (m)	Z (米)
1	500	500	80	100	6	245	458	30	89
2	200	200	15	70	7	660	150	40	56
3	900	567	25	85	8	900	328	22	78
4	850	820	35	80	9	326	895	17	78
5	150	698	18	60	10	628	736	20	91

Table 3. Parameter value.

表 3. 参数值。

Parameter	Value
Total number of training episodes	5000
Discount factor	0.99
Target network update rate	0.001
Buffer size	10,000
Batch size	100
The initial destination area radius: R	10
Attenuation coefficient: C	0.8

参数	值
训练阶段的总数	5000
折扣因子	0.99
目标网络更新率	0.001
缓冲区大小	10,000
批次大小	100
初始目标区域半径: R	10
衰减系数: C	0.8



(a) (b)

Figure 6. The experimental environment. (a) Two-dimensional perspective. (b) Three-dimensional perspective.

图 6. 实验环境。(a) 二维视角。(b) 三维视角。

4.2. Collision Avoidance Agent Training

4.2. 避障代理训练

For various conflicts arising from UAV pairs, the trained agent can provide appropriate solutions. During training, the average reward obtained per episode is an important indicator of convergence and collision avoidance performance. Using the improved DDPG algorithm to train agents, the rewards for each episode are shown in Figure 7.

对于由无人机对产生的各种冲突，训练好的代理能够提供适当的解决方案。在训练过程中，每轮获得的平均奖励是收敛性和避障性能的重要指标。使用改进的 DDPG 算法训练代理，每轮的奖励如图 7 所示。

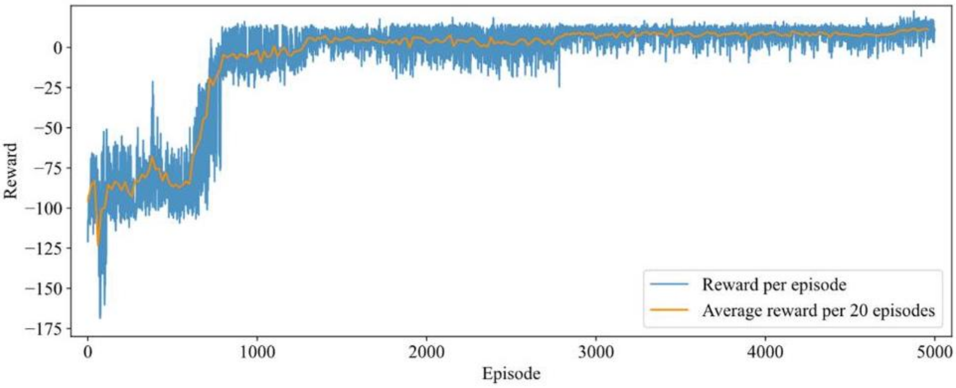


Figure 7. The rewards for each episode during training.

图 7. 训练期间每轮的奖励。

From Figure 7, the reward obtained by the agent is not stable at the beginning of the training, as the agent touches the events with a higher degree of punishment during the exploration process, resulting in a large degree of reward drop. With continuous training, the agent gradually learns the high-reward behavior, and the reward value increases. In the second half of training, the reward did not significantly fall again, which indicates that the improved algorithm learned a better and more stable strategy, and therefore the reward oscillated less.

从图 7 可以看出，训练开始时代理获得的奖励并不稳定，因为代理在探索过程中触发了惩罚程度较高的事件，导致奖励大幅下降。随着训练的持续进行，代理逐渐学会了高奖励行为，奖励值增加。在训练的后半段，奖励没有再次显著下降，这表明改进的算法学会了更好且更稳定的策略，因此奖励波动较小。

The comparison effect of the improved DDPG algorithm with the original algorithm is shown in Figure 8. It can be seen that the improved algorithm demonstrates a better improvement in convergence speed, achieving a higher reward value and showing a convergence trend of around 380 training epochs, while the original algorithm was only able to show such an effect after around 1000 epochs. After the 2000th training epoch, the rewards obtained by the two algorithms did not differ much. However, as for actual training, the improved DDPG algorithm obtained stable reward values and determined the convergence trend at earlier epochs, and thus training can be ended earlier than in the original algorithm, which saves training time.

改进的 DDPG 算法与原算法的比较效果如图 8 所示。可以看出，改进的算法在收敛速度上表现出更好的提升，实现了更高的奖励值，并在大约 380 个训练周期时显示出收敛趋势，而原算法仅在大约 1000 个周期后才能显示出这样的效果。在第 2000 个训练周期之后，两种算法获得的奖励差异不大。然而，在实际训练中，改进的 DDPG 算法获得了稳定的奖励值，并在较早的周期确定了收敛趋势，因此训练可以比原算法更早结束，从而节省了训练时间。

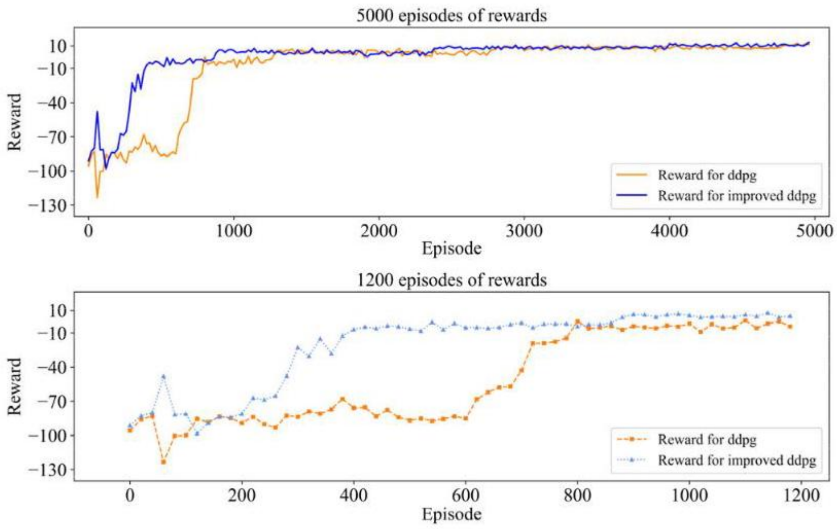


Figure 8. Comparison of the convergence process.
图 8. 收敛过程的比较。

4.3. Numerical Results Analysis

4.3. 数值结果分析

4.3.1. Collision Avoidance Results

4.3.1. 避碰结果

Using the two-layer resolution framework, we obtained the collision avoidance results, as shown in Figure 9, while recording the distance between each UAV and the nearest obstacle, as well as the distance between the two nearest UAVs.

使用双层分辨率框架，我们获得了如图 9 所示的避碰结果，同时记录了每个无人机与最近障碍物之间的距离以及两个最近无人机之间的距离。

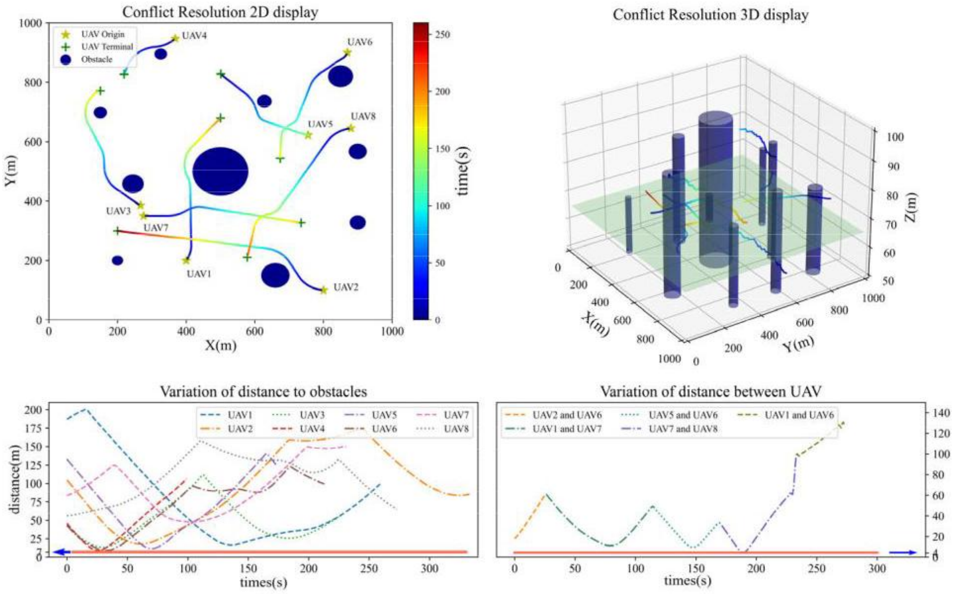


Figure 9. Collision avoidance results.

图 9. 避碰结果。

From a two-dimensional perspective, it is intuitively apparent that every UAV is in conflict with at least one obstacle and avoids obstacles with as little extra flight distance as possible. In addition, there is a risk of conflict between UAV5 and UAV6, UAV7 and UAV8, and UAV1 and UAV7, so the agent randomly selects one of the UAVs to perform the primary avoidance maneuver, while the other maintains almost its original direction (or makes minor adjustments), in order to minimize the impact of the avoidance behavior on normal navigation. In the three-dimensional view, all the UAVs have reached the intended altitude.

从二维视角来看，直观上显然每个无人机至少与一个障碍物存在冲突，并且尽可能地以最少的额外飞行距离避开障碍物。此外，UAV5 与 UAV6、UAV7 与 UAV8 以及 UAV1 与 UAV7 之间存在冲突风险，因此代理随机选择其中一个无人机执行主要避障机动，而另一个几乎保持原有方向 (或进行小幅度调整)，以最小化避障行为对正常导航的影响。在三维视角中，所有无人机都已达到预期的高度。

The distance between each UAV and the obstacle has a process from small to large, which indicates that UAVs are making avoidance actions. The closest UAV pairs may be different at different times, but the overall trend of distance variation is consistent, proving that the UAVs can also avoid each other. The minimum distance from buildings is about 7 m , and the minimum distance from other UAVs is about 4 m throughout the whole process, thus meeting the standard safety interval, proving that the model in this paper can ensure the safe operation of UAVs in cities with many buildings.

每个无人机与障碍物之间的距离有一个从小到大的过程，这表明无人机正在进行避障动作。最近的无人机对在不同时间可能有所不同，但距离变化的总体趋势是一致的，证明无人机也能相互避让。整个过程中，无人机与建筑物的最小距离约为 7 m ，与其他无人机的最小距离约为 4 m ，从而满足了标准安全间隔，证明了本文中的模型能够确保无人机在多建筑城市中的安全运行。

4.3.2. Avoidance Strategy Analysis

4.3.2. 避障策略分析

In the two-layer resolution framework, three strategies are used for collision avoidance and destination guidance, to analyze the avoidance action selection pattern by the agent, recording the actions (heading angle, altitude, speed change) selected by all UAVs at each step, as shown in Figures 10-12.

在双层分辨率框架中，采用了三种策略来进行碰撞避免和目的地引导，以分析代理的避障动作选择模式，记录了如图 10-12 所示的在每一步所有无人机选择的动作 (航向角、高度、速度变化)。

As shown in Figure 10, each approach of UAVs and obstacles will lead to a significant change in the heading angle, and when the distance is kept at a relatively safe level, the change in heading angle will fluctuate around 0° , indicating that the UAV is flying along a straight line in the horizontal direction. It can be determined that the agent avoids collision with obstacles mainly by changing the heading angle of the UAV.

如图 10 所示，无人机与障碍物的每一次接近都会导致航向角发生显著变化，并且当距离保持在相对安全的水平时，航向角的变化将在 0° 左右波动，表明无人机在水平方向上沿直线飞行。可以确定，代理主要通过改变无人机的航向角来避免与障碍物碰撞。

As shown in Figures 11 and 12, climbing and descending actions ensure that the height of the UAV is finally consistent with the destination height, demonstrating that the agent has the guiding ability in the three-dimensional space. The speed change is generally stable within a fixed range, and there is no excessive speed, as the speed adjustment is coupled with the heading angle to avoid collision with obstacles. In addition, due to there being fewer obstacles near the destination, the UAV has a higher speed and a stable heading angle in the later stage, ultimately reaching the destination.

如图 11 和 12 所示，爬升和下降动作确保了无人机的最终高度与目的地高度一致，表明代理在三维空间中具有引导能力。速度变化通常在固定范围内保持稳定，没有过度的速度，因为速度调整与航向角相结合以避免与障碍物碰撞。此外，由于目的地附近障碍物较少，无人机在后期阶段速度较高，航向角稳定，最终到达目的地。

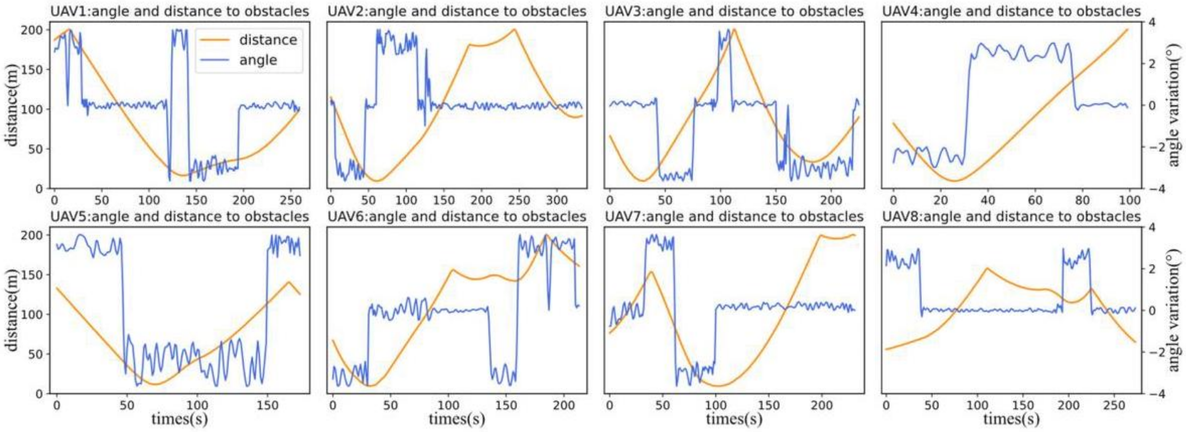


Figure 10. The heading angle change trend and distance to obstacles of UAV.
图 10. 无人机的航向角变化趋势和与障碍物的距离。
Figure 12. The speed change trend of UAVs.
图 12. 无人机的速度变化趋势。

4.4. Performance Analysis

4.4. 性能分析

4.4.1. Performance Testing of the Method

4.4.1. 方法的性能测试

To fully illustrate the collision avoidance effect of the method, 300 different scenarios were designed in the simulation area by randomly generating the starting and ending points of eight UAVs. The considered performance metrics were the following: (1) collision avoidance success rate (SR), which is the percentage of successful collision avoidance; (2) computational efficiency (CE), which is the time required for an agent to calculate an action; (3) extra flight distance (ED), which is the record of extra distance the UAV flighted due to collision avoidance.

为了充分说明该方法避障效果，在仿真区域内随机生成八架无人机的起点和终点，设计了 300 个不同场景。考虑的性能指标包括以下几项:(1) 避障成功率 (SR)，即成功避障的百分比；(2) 计算效率 (CE)，即代理计算一个动作所需的时间；(3) 额外飞行距离 (ED)，即记录无人机因避障而飞行的额外距离。

(1) Collision avoidance success rate

(1) 避障成功率

In the 300 random scenarios, 4523 conflicts were recorded in the conflict resolution pool, including 867 conflicts with other UAVs and 3656 conflicts with buildings. If the agent cannot assign the correct avoidance action to a UAV, a collision will occur, according to the current speed trend. The collision avoidance method described in this paper can guide UAVs out of collision risk, with success rates as shown in Table 4.

在 300 个随机场景中，冲突解决池记录了 4523 次冲突，其中包括与其他无人机冲突的 867 次和与建筑物的 3656 次冲突。如果代理无法为无人机分配正确的避障动作，根据当前速度趋势，将会发生碰撞。本文描述的避障方法可以引导无人机摆脱碰撞风险，成功率如表 4 所示。

Table 4. Success rates of collision avoidance.
表 4. 避障成功率。

	With Buildings	With Other UAVs	Total
Number of conflicts	3656	867	4523
Number of resolutions	3502	796	4298
Success rate	95.79%	91.81%	95.03%

	与建筑物	与其他无人机	总计
冲突数量	3656	867	4523
解决数量	3502	796	4298
成功率	95.79%	91.81%	95.03%

The data in Table 4 show that the method has a higher success rate when resolving conflicts with fixed obstacles than with dynamic obstacles, as the invading UAV has positional uncertainty, the flight state may not be fully perceived by the current UAV, and no avoidance action can be taken in time. However, the overall success rate reached 95.03%, indicating that the method is able to guide UAVs to avoid most collision risks and can provide an adequate and reliable reference for urban air traffic management.

表 4 中的数据表明，该方法在解决与固定障碍物的冲突时成功率较高，而在与动态障碍物的冲突中成功率较低，因为入侵的无人机存在位置不确定性，当前无人机可能无法完全感知飞行状态，并且无法及时采取避障动作。然而，整体成功率达到了 95.03%，表明该方法能够引导无人机避免大多数碰撞风险，并为城市空中交通管理提供充足可靠的参考。

(2) Computational efficiency

(2) 计算效率

Recording the total computation time and the number of avoidance actions performed by UAVs in each scenario, to calculate the average time that the method to plan an action for a UAV, and this is used as the evaluation index for computation efficiency, as shown in Equation (24).

记录每个场景中无人机的总计算时间和避障动作次数，以计算为无人机规划动作的平均时间，这作为计算效率的评价指标，如公式 (24) 所示。

$$T_f^j = \frac{T_{\text{total}}^j}{\sum\limits_{i=1}^{8} \text{num}_i^j} \tag{24}$$

where T_{total}^j is the total time required to calculate an avoidance action in scenario j , and num_i^j is the number of avoidance actions performed by UAV i .

其中 T_{total}^j 是在场景 j 中计算避障动作所需的总时间，num_i^j 是无人机 i 执行的避障动作次数。

The average time required for each scenario is shown in Figure 13. From the figure, the avoidance action calculation time of 300 scenarios is at the 0.01 s level, with an average time of 0.0963 s , which can meet the real-time requirements of collision avoidance.

每个场景所需的平均时间如图 13 所示。从图中可以看出，300 个场景的避障动作计算时间处于 0.01 s 水平，平均时间为 0.0963 s ，可以满足避障的实时性要求。

(3) Extra flight distance

(3) 额外飞行距离

When facing obstacles, the agent guides the UAV to change its heading or speed, which adds extra flight distance (ED) compared with the original trajectory. The metrics of ED are used to measure the impact of collision avoidance on UAVs, as shown in Equation (25):

当面对障碍物时，代理引导无人机改变航向或速度，与原始轨迹相比增加了额外的飞行距离 (ED)。ED 的指标用于衡量避障对无人机的影响，如公式 (25) 所示:

$$d_e = \frac{1}{n} \sum_{i=1}^{n} (d_{oi} - d_{ni}) \tag{25}$$

In Equation (25), d_e means the extra flight distance (ED), d_{oi} means the distance of the i -th trajectory directly to the destination regardless of any conflicts, d_{ni} means the distance of the i -th trajectory which has considered the collision avoidance. n is the total number of trajectories in all scenarios.

在公式 (25) 中，d_e 表示额外的飞行距离 (ED)，d_{oi} 表示不考虑任何冲突直接到达目的地的第 i 条轨迹的距离，d_{ni} 表示考虑了避障的第 i 条轨迹的距离。n 是所有场景中轨迹的总数。

From the perspective of flight efficiency and green transportation, the shorter the ED, the less flight energy is lost, and the less impact there is on the original flight [29]. The average extra flight distance of eight aircraft in300scenarios is 26.8 m , which is relatively good and is acceptable in terms of a collaborative resolution process for mUAVs.

从飞行效率和绿色交通的角度来看，ED 越短，飞行中损失的能源越少，对原始飞行的影响也越小 [29]。在 300 个场景中，八架飞机的平均额外飞行距离为 26.8 m ，在多无人机协作解决过程中相对较好，是可接受的。

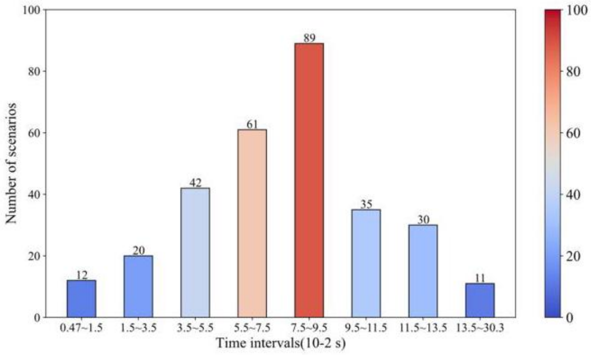


Figure 13. The average time required for allocating actions in each scenario.

图 13. 分配每个场景中动作所需的平均时间。

4.5. Impact of Noisy States

4.5. 噪声状态的影响

UAVs may have positional uncertainty due to the interference of random factors such as crosswinds, which can affect collision avoidance behavior. To investigate the robustness of the proposed method, we performed simulations with noisy states, adding noise to the state information of UAV. It was assumed that each noise component was uniformly distributed, i.e., $m=\left[m_x, m_y, m_z, m_v\right], m \sim U(-\varepsilon, \varepsilon)$. The noise was added to the state information of UAV, i.e., $P=\left[x_i^t+m_x, y_i^t+m_y, z_i^t+m_z\right]$ and $\bar{V}_i^t=V_i^t+m_v$, which will change the state of the agent in Equation (4).

由于随机因素如侧风的干扰，无人机可能存在位置不确定性，这可能会影响避障行为。为了研究所提方法的鲁棒性，我们在带有噪声的状态下进行了模拟，向无人机状态信息中添加了噪声。假设每个噪声分量都是均匀分布的，即 $m=\left[m_x, m_y, m_z, m_v\right], m \sim U(-\varepsilon, \varepsilon)$ 。噪声被添加到无人机状态信息中，即 $P=\left[x_i^t+m_x, y_i^t+m_y, z_i^t+m_z\right]$ 和 $\bar{V}_i^t=V_i^t+m_v$ ，这将改变方程 (4) 中代理的状态。

Table 5 shows the SR and ED performances with noisy states. The table shows that noise influences the SR and ED performance, meaning that the UAV does not have accurate position and velocity information, leading to biased observations by the agent, which may output incorrect avoidance actions. However, even if noise $\varepsilon=3$ is added to both position and velocity information, SR is, at most, 91.61% , and ED is 36.2 m , and thus a good level is still maintained. Therefore, our method has high tolerance to noisy observations.

表 5 显示了带有噪声状态下的 SR 和 ED 性能。表格显示噪声影响了 SR 和 ED 性能，意味着无人机没有准确的位置和速度信息，导致代理产生偏见的观测，这可能会输出错误的避障动作。然而，即使向位置和速度信息中添加了噪声 $\varepsilon=3$ ，SR 最多为 91.61% ，ED 为 36.2 m ，因此仍然保持了良好的水平。因此，我们的方法对噪声观测具有高容忍度。

Table 5. SR, and ED performance when states are noisy.

表 5。当状态存在噪声时，SR 和 ED 的性能。

	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	No Noise
	$vx = 0, \varepsilon = 0.5$	$vx = 0, \varepsilon = 1.5$	$vx = 0, \varepsilon = 3$	
SR	94.32%	93.98%	92.90%	95.03%
ED	27.2 m	30.3 m	35.6 m	26.8 m
	$mx, mx, mx = 0$	$mx, mx, mx = 0$	$mx, mx, mx = 0$	
	$vx \neq 0, \varepsilon = 0.5$	$vx \neq 0, \varepsilon = 1.5$	$vx \neq 0, \varepsilon = 3$	
SR	94.54%	93.52%	93.3%	
ED	27.5 m	29.6 m	34.3 m	
	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	
	$vx \neq 0, \varepsilon = 0.5$	$vx \neq 0, \varepsilon = 1.5$	$vx \neq 0, \varepsilon = 3$	
SR	94.01%	92.83%	91.61%	
ED	27.6 m	33.3 m	36.2 m	

	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	无干扰
	$vx = 0, \varepsilon = 0.5$	$vx = 0, \varepsilon = 1.5$	$vx = 0, \varepsilon = 3$	
成功率 (SR)	94.32%	93.98%	92.90%	95.03%
预测误差 (ED)	27.2 m	30.3 m	35.6 m	26.8 m
	$mx, mx, mx = 0$	$mx, mx, mx = 0$	$mx, mx, mx = 0$	
	$vx \neq 0, \varepsilon = 0.5$	$vx \neq 0, \varepsilon = 1.5$	$vx \neq 0, \varepsilon = 3$	
成功率 (SR)	94.54%	93.52%	93.3%	
预测误差 (ED)	27.5 m	29.6 m	34.3 m	
	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	$mx, mx, mx \neq 0$	
	$vx \neq 0, \varepsilon = 0.5$	$vx \neq 0, \varepsilon = 1.5$	$vx \neq 0, \varepsilon = 3$	
成功率 (SR)	94.01%	92.83%	91.61%	
预测误差 (ED)	27.6 m	33.3 m	36.2 m	

4.6. Different Numbers of UAVs

4.6. 无人机数量的不同

Table 6 presents the CR, SR, and DR performances in scenarios with different numbers of UAVs J $\in \{2, 4, 8, 10, 12, 20\}$. The rates are averaged over 300 random realizations (all UAVs having random starting points and destinations). From the table, it can be noted that with increasing numbers of UAVs, the SR decreases due to the higher risk of collision, the ED has an overall upward trend, while the CE is not affected by the number of UAVs.

表 6 展示了在不同数量的无人机场景下的 CR、SR 和 DR 性能 J $\in \{2, 4, 8, 10, 12, 20\}$ 。这些比率是在 300 次随机实现 (所有无人机具有随机起点和终点) 上平均得出的。从表中可以看出，随着无人机数量的增加，由于碰撞风险的增加，SR 降低，ED 总体上呈上升趋势，而 CE 不受无人机数量的影响。

Table 6. SR, CE, and ED performance with different numbers of UAVs.

表 6。不同数量无人机的 SR、CE 和 ED 性能。

	$J = 2$	$\mathbf{J = 4}$	$J = 8$	$\mathbf{J = 10}$	$\mathbf{J = 12}$	$J = 20$
SR	100%	99.1%	95.03%	95.62%	92.10%	90.56%
CE	0.0832 s	0.0721 s	0.0963 s	0.1861 s	0.0910 s	0.216 s
ED	20.3 m	25.2 m	26.8 m	27.6 m	39.1 m	38.2 m

	$J = 2$	$\mathbf{J = 4}$	$J = 8$	$\mathbf{J = 10}$	$\mathbf{J = 12}$	$J = 20$
成功率 (SR)	100%	99.1%	95.03%	95.62%	92.10%	90.56%
通信效率 (CE)	0.0832 s	0.0721 s	0.0963 s	0.1861 s	0.0910 s	0.216 s
ED	20.3 m	25.2 m	26.8 m	27.6 m	39.1 m	38.2 m

We note that when there are 20 UAVs in an urban scenario of 1 square kilometer, our method still maintains a success rate of more than 90% in terms of conflict, indicating that this method can provide safe guidance for 20 UAVs in that area, which is sufficient to adapt to the current scale of urban UAVs.

我们注意到，在 1 平方公里的城市环境中，当有 20 架无人机时，我们的方法在冲突方面的成功率仍超过 90% ，这表明这种方法可以为该区域内的 20 架无人机提供安全指导，足以适应当前城市无人机的规模。

In Figure 14, illustrations of collision avoidance in scenarios with different numbers of UAVs J $\in \{2, 8, 20\}$ are presented. From Figure 14, it can be observed that due to the different locations and numbers of UAVs, the agent may determine different avoidance actions when facing collision risks, leading to different trajectories.

在图 14 中，展示了不同数量的无人机 J $\in \{2,8,20\}$ 环境下的避障示例。从图 14 中可以观察到，由于无人机的位置和数量不同，当面临碰撞风险时，代理可能会确定不同的避障行为，导致不同的轨迹。

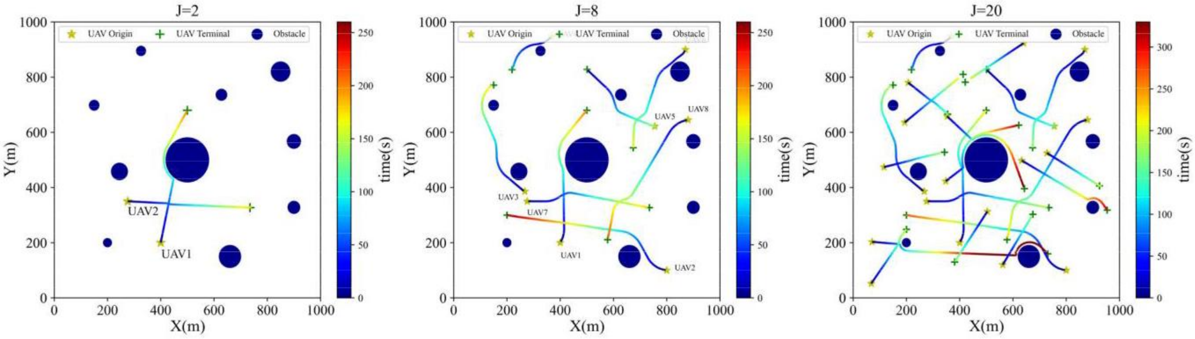


Figure 14. Illustrations of collision avoidance in different scenarios that J $\in \{2,8,20\}$.

图 14. 不同场景下的避障示例 J $\in \{2,8,20\}$ 。

4.7. Comparison with Other Algorithms

4.7. 与其他算法的比较

In the same scenario, two other algorithms are selected for comparison: the double deep Q network (DDQN) and the artificial potential field (APF).

在同一场景中，选择了两种其他算法进行比较: 双深度 Q 网络 (DDQN) 和人工势场 (APF)。

DDQN is a typical value-based DRL algorithm [18,19], while APF utilizes the repulsive force of obstacles and the gravitational force of target to guide the UAV motion, and is widely used in research on collision avoidance [30,31]. DDQN requires that the agent action space be discrete, which is set to $\Delta\varphi \in \{-3^\circ, 0, 3^\circ\}, \Delta Z \in \{-1\text{ m}, 0\text{ m}, 1\text{ m}\}$, $\Delta V \in \{-2\text{ m/s}, 0\text{ m/s}, 2\text{ m/s}\}$. The APF action space is consistent with our method. Experiments using DDQN were conducted based on a two-layer resolution framework, and the APF was divided into two categories: APF with a two-layer framework and APF without a two-layer framework. Table 7 presents the performances of the different algorithms.

DDQN 是一种典型的基于价值的深度强化学习算法 [18,19]，而 APF 利用障碍物的排斥力和目标的引力来引导无人机运动，在避障研究 [30,31] 中被广泛应用。DDQN 要求代理动作空间是离散的，设置为 $\Delta\varphi \in \{-3^\circ, 0, 3^\circ\}, \Delta Z \in \{-1\text{ m}, 0\text{ m}, 1\text{ m}\}$ ，$\Delta V \in \{-2\text{ m/s}, 0\text{ m/s}, 2\text{ m/s}\}$ 。APF 的动作空间与我们的方法一致。基于两层分辨率框架进行的 DDQN 实验，APF 被分为两类: 具有两层框架的 APF 和没有两层框架的 APF。表 7 展示了不同算法的性能。

Table 7. SR, CE, and ED performance of different algorithms and J = 8 .

表 7. 不同算法的 SR、CE 和 ED 性能以及 J = 8。

	SR	CE	ED
APF without two-layer framework	89.32%	1.8329 s	41.2 m
APF with two-layer framework	91.65%	0.0821 s	34.3 m
DDQN	93.83%	0.1839 s	56.3 m
Improved DDPG	95.03%	0.0963 s	26.8 m

	SR	CE	ED
无两层框架的 APF	89.32%	1.8329 s	41.2 m
有两层框架的 APF	91.65%	0.0821 s	34.3 m
双深度 Q 网络 (DDQN)	93.83%	0.1839 s	56.3 m
改进的深度确定性策略梯度 (DDPG)	95.03%	0.0963 s	26.8 m

The improved DDPG algorithm has an absolute advantage in terms of the SR, which is much greater than the other algorithms, at 95.03% . There was not much difference in CE, and only APF was higher than the other algorithms. Our method achieved the minimum ED, which was the largest value in

DDQN; because of the constraints of the discrete action space, the DDQN-trained agent can only perform a limited number of action values and the flexibility of the UAV cannot be fully utilized. In addition, the performance of the APF with a two-layer framework was superior to that of the original APF algorithm, indicating that the two-layer resolution framework can better avoid conflict risks when presented with mUAVs.

改进的 DDPG 算法在 SR 方面具有绝对优势，该优势远大于其他算法，在 95.03% 处。CE 方面的差异不大，只有 APF 高于其他算法。我们的方法实现了最小的 ED，这是 DDQN 中的最大值；由于离散动作空间的限制，DDQN 训练的智能体只能执行有限数量的动作值，无法充分利用无人机的灵活性。此外，双层框架下的 APF 性能优于原始 APF 算法，表明双层解析框架在面临多无人机时能更好地避免冲突风险。

It is worth noting that we observed in training sessions that the improved DDPG completed convergence faster than the DDQN, which is consistent with our results in Section 4.2, compared to the original DDPG algorithm, and further indicates that the dynamic adjustment mechanism can reduce the number of training episodes.

值得注意的是，我们在训练过程中观察到改进的 DDPG 比 DDQN 更快地完成收敛，这与我们在第 4.2 节中的结果一致，与原始 DDPG 算法相比，进一步表明动态调整机制可以减少训练环节的数量。

5. Conclusions

5. 结论

In this paper, an adaptive method for mUAV collision avoidance in urban air traffic was studied. The main conclusions are as follows:

在本文中，研究了城市空中交通中多无人机避障的自适应方法。主要结论如下:

Firstly, the proposed two-layer resolution framework provides a new concept for realizing mUAV collision avoidance, in which each UAV is endowed with decision-making ability, and the computational complexity is controlled at the polynomial level. Using the improved DDPG algorithm to train the agent allows convergence to be completed faster, which saves training costs to a great extent.

首先，提出的双层解析框架为实现在城市空中交通中多无人机避障提供了新概念，在该框架中，每个无人机都被赋予决策能力，并且计算复杂度控制在多项式级别。使用改进的 DDPG 算法训练智能体可以更快地完成收敛，这在很大程度上节省了训练成本。

Secondly, the numerical results indicate that the proposed method is able to adapt to various scenarios, e.g., different numbers and positions of UAVs, and interference from random factors. More specifically, the average decision time of the method is 0.0963 s with eight UAVs, the overall resolution success rate is 95.03% , and the extra flight distance is 26.8 m . Our method has better performance when compared to APF, APF with a two-layer framework, and DDQN.

其次，数值结果表明，提出的方法能够适应各种场景，例如，不同数量和位置的无人机以及随机因素的干扰。更具体地说，该方法在八架无人机情况下的平均决策时间为 0.0963 s ，整体分辨率成功率为 95.03% ，额外飞行距离为 26.8 m 。与 APF、带有两层框架的 APF 和 DDQN 相比，我们的方法具有更好的性能。

Thirdly, from the perspective of the avoidance process, changing the heading angle is the main way of avoiding collision, the minimum distance from buildings is about 7 m , and the minimum distance from other UAVs is about 4 m , which further proves that the method has a relatively high sensitivity for static obstacles. Our future research focus will be on how to determine the appropriate safety interval and how to reflect this in the resolution process.

第三，从避障过程的角度来看，改变航向角是避免碰撞的主要方式，与建筑物的最小距离约为 7 m ，与其他无人机的最小距离约为 4 m ，这进一步证明了该方法对静态障碍物具有较高的敏感性。我们未来的研究重点将是确定适当的安全间隔以及如何在解决过程中体现这一点。

Despite the strengths of our proposed approach, there are some drawbacks that require further study. In this paper, distance and velocity vector were calculated for conflict detection, which lacks objective quantification of conflict risk and may have an impact on the subsequent collision avoidance. The quantitative assessment of UAV conflict risk based on multiple factors could guarantee a more accurate determination of conflict targets and resolution strategies, which would be a valuable research direction in the future. Another valuable research direction would be to combine kinematics theory and control theory. Assigning appropriate resolution strategies for UAV collision avoidance at the level of urban air traffic management, while designing UAV controllers from the perspective of control performance, thus ensuring that UAVs can successfully complete avoidance actions by formulating suitable control parameters and suppressing the influence of external disturbances [32,33]. This would promote the engineering application of the method described in this paper.

尽管我们提出的方法具有一定的优势，但仍存在需要进一步研究的不足之处。在本文中，为了冲突检测，计算了距离和速度向量，这缺乏对冲突风险的客观量化，可能会对后续的避障产生影响。基于多个因素的无人机冲突风险定量评估可以确保更准确地确定冲突目标和解决策略，这将是未来的一个有价值的研究方向。另一个有价值的研究方向将是结合动力学理论和控制理论。在城区空中交通管理层面为无人机避障分配适当的解决策略，同时从控制性能的角度设计无人机控制器，从而通过制定合适的控制参数和抑制外部干扰，确保无人机能够成功完成避障动作。这将促进本文描述的方法的工程应用。

Author Contributions: Conceptualization, H.Z.; methodology, J.Z. (Jinpeng Zhang) and H.Z.; software, J.Z. (Jinpeng Zhang) and J.Z. (Jinlun Zhou); validation, J.Z. (Jinpeng Zhang), M.H. and G.Z.; formal analysis, J.Z. (Jinpeng Zhang) and H.L.; investigation, M.H. and J.Z. (Jinlun Zhou); resources, J.Z. (Jinpeng Zhang), H.Z. and H.L.; writing-original draft preparation, J.Z. (Jinpeng Zhang); writing-review and editing, J.Z. (Jinpeng Zhang) and M.H.; visualization, J.Z. (Jinpeng Zhang) and M.H. All authors have read and agreed to the published version of the manuscript.

作者贡献: 构思，H.Z.；方法论，J.Z. (Jinpeng Zhang) 和 H.Z.；软件，J.Z. (Jinpeng Zhang) 和 J.Z. (Jinlun Zhou)；验证，J.Z. (Jinpeng Zhang), M.H. 和 G.Z.；形式分析，J.Z. (Jinpeng Zhang) 和 H.L.；调查研究，M.H. 和 J.Z. (Jinlun Zhou)；资源，J.Z. (Jinpeng Zhang), H.Z. 和 H.L.；写作-原始草稿准备，J.Z. (Jinpeng Zhang)；写作-审查与编辑，J.Z. (Jinpeng Zhang) 和 M.H.；可视化，J.Z. (Jinpeng Zhang) 和 M.H. 所有作者已阅读并同意发表的稿件版本。

Funding: This research was funded by National Natural Science Foundation of China (71971114).

资金: 本研究得到了中国国家自然科学基金 (71971114) 的资助。

Data Availability Statement: Data will be made available on request.

数据可获取性声明: 数据将在请求时提供。

Abbreviations

缩写

D_u, D_o The collision zone of UAV and building

D_u, D_o 无人机与建筑的碰撞区域

The heading angle, horizontal speed and latitude

航向角、水平速度和纬度

The horizontal speed of UAV

无人机的水平速度

Z The latitude of UAV

Z 无人机的纬度

φ_g The relative heading angle of the destination to UAV

φ_g 目标相对于无人机的航向角

d_g The horizontal distance of the destination to UAV

d_g 目标相对于无人机的水平距离

φ_{us} The difference in heading angle between two UA

φ_{us} 两个无人机之间的航向角差

Z_{us} The difference in altitude between two UAV

Z_{us} 两个无人机之间的高度差

d_{us} The horizontal distance between two UAV

d_{us} 两个无人机之间的水平距离

d_{det} The UAV detection distance

d_{det} 无人机的检测距离

$\pi(s;\theta)$ Policy network

$\pi(s;\theta)$ 策略网络

$\pi(s;\theta^-)$ Target policy network

$\pi(s;\theta^-)$ 目标策略网络

$\pi_\theta(s_i)$ The learned policy function

$\pi_\theta(s_i)$ 学到的策略函数

i, j Superscript or subscript, denote the specific UAV

i, j 上标或下标，表示特定的无人机

$\Delta\varphi$ Alteration in direction

$\Delta\varphi$ 方向的改变

ΔZ Alteration in altitude

ΔZ 高度的改变

ΔV Alteration in horizontal speed

ΔV 水平速度的改变

$d_{\text{ om}}$ The distance attribute of obstacle in sector m

$d_{\text{ om}}$ 扇区 m 中障碍物的距离属性

R_d Destination intent or collision avoidance rewar

R_d 目的地意图或避障奖励

R_{uu} The UAV collision avoidance reward

R_{uu} 无人机避障奖励

R_{ex} The additional reward

R_{ex} 额外奖励

S The state space of agent

S 代理的状态空间

A The action space of agent

A 代理的动作空间

P The collision resolution pool

P 碰撞解决池

$Q(s, a; \omega)$ Q network

$Q(s, a; \omega)$ Q 网络

$Q(s, a; \omega^-)$ Target Q network

$Q(s, a; \omega^-)$ 目标 Q 网络

$Q_\omega(s, a)$ The learned value function

$Q_\omega(s, a)$ 学习到的价值函数

t Subscript, denote the specific moment

t 下标，表示特定时刻

References

参考文献

1. Garrow, L.A.; German, B.J.; Leonard, C.E. Urban air mobility: A comprehensive review and comparative analysis with autonomous and electric ground transportation for informing future research. Transp. Res. Part C Emerg. Technol. 2021, 132, 103377. [CrossRef]

2. Barrado, C.; Boyero, M.; Brucculeri, L.; Ferrara, G.; Hately, A.; Hullah, P.; Martin-Marrero, D.; Pastor, E.; Rushton, A.P.; Volkert, A. U-Space Concept of Operations: A Key Enabler for Opening Airspace to Emerging Low-Altitude Operations. Aerospace 2020, 7, 24. [CrossRef]

3. 2022 Civil Aviation Development Statistical Bulletin. 2023. Available online: https://file.veryzhun.com/buckets/car 7390295f32633128e6e5cee44fc9fe4e.pdf (accessed on 1 May 2023).

4. Kiran, B.R.; Sobh, I.; Talpaert, V.; Mannion, P.; Sallab, A.A.A.; Yogamani, S.; Pérez, P. Deep Reinforcement Learning for Autonomous Driving: A Survey. IEEE Trans. Intell. Transp. Syst. 2022, 23, 4909-4926. [CrossRef]

5. Wu, Y. A survey on population-based meta-heuristic algorithms for motion planning of aircraft. Swarm Evol. Comput. 2021, 62, 100844. [CrossRef]

6. Zeng, D.; Chen, H.; Yu, Y.; Hu, Y.; Deng, Z.; Leng, B.; Xiong, L.; Sun, Z. UGV Parking Planning Based on Swarm Optimization and Improved CBS in High-Density Scenarios for Innovative Urban Mobility. Drones 2023, 7, 295. [CrossRef]

7. Zhao, P.; Erzberger, H.; Liu, Y. Multiple-Aircraft-Conflict Resolution Under Uncertainties. J. Guid. Control Dyn. 2021, 44, 2031-2049. [CrossRef]

8. Yun, S.C.; Ganapathy, V.; Chien, T.W. Enhanced D* Lite Algorithm for mobile robot navigation. In Proceedings of the 2010 IEEE Symposium on Industrial Electronics and Applications (ISIEA), Penang, Malaysia, 3-5 October 2010; pp. 545-550.

9. Wu, Y.; Low, K.H.; Pang, B.; Tan, Q. Swarm-Based 4D Path Planning For Drone Operations in Urban Environments. IEEE Trans Veh. Technol. 2021, 70, 7464-7479. [CrossRef]

10. Zhang, Q.; Wang, Z.; Zhang, H.; Jiang, C.; Hu, M. SMILO-VTAC Model Based Multi-Aircraft Conflict Resolution Method in Complex Low-Altitude Airspace. J. Traffic Transp. Eng. 2019, 19, 125-136.

11. Radmanesh, M.; Kumar, M. Flight formation of UAVs in presence of moving obstacles using fast-dynamic mixed integer linear programming. Aerosp. Sci. Technol. 2016, 50, 149-160. [CrossRef]

12. Waen, J.D.; Dinh, H.T.; Torres, M.H.C.; Holvoet, T. Scalable multirotor UAV trajectory planning using mixed integer linear programming. In Proceedings of the 2017 European Conference on Mobile Robots (ECMR), Paris, France, 6-8 September 2017; pp. 1-6.

13. Alonso-Ayuso, A.; Escudero, L.F.; Martín-Campo, F.J. An exact multi-objective mixed integer nonlinear optimization approach for aircraft conflict resolution. TOP 2016, 24, 381-408. [CrossRef]

14. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. Nature 2015, 518, 529-533. [CrossRef] [PubMed]

15. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016, 529, 484-489. [CrossRef]

16. Pham, H.X.; La, H.M.; Feil-Seifer, D.; Nguyen, L.V. Autonomous uav navigation using reinforcement learning. arXiv 2018, arXiv:1801.05086.

17. Liu, X.; Liu, Y.; Chen, Y. Reinforcement Learning in Multiple-UAV Networks: Deployment and Movement Design. IEEE Trans. Veh. Technol. 2019, 68, 8036-8049. [CrossRef]

18. Singla, A.; Padakandla, S.; Bhatnagar, S. Memory-Based Deep Reinforcement Learning for Obstacle Avoidance in UAV with Limited Environment Knowledge. IEEE Trans. Intell. Transp. Syst. 2021, 22, 107-118. [CrossRef]

19. Zhai, P.; Zhang, Y.; Shaobo, W. Intelligent Ship Collision Avoidance Algorithm Based on DDQN with Prioritized Experience Replay under COLREGs. J. Mar. Sci. Eng. 2022, 10, 585. [CrossRef]

20. Li, C.; Gu, W.; Zheng, Y.; Huang, L.; Zhang, X. An ETA-Based Tactical Conflict Resolution Method for Air Logistics Transportation. Drones 2023, 7, 334. [CrossRef]

21. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. arXiv 2015, arXiv:1509.02971.

22. Ribeiro, M.; Ellerbroek, J.; Hoekstra, J. Improvement of conflict detection and resolution at high densities through reinforcement learning. In Proceedings of the ICRAT2020: International Conference on Research in Air Transportation 2020, Tampa, FL, USA, 23-26 June 2020.

23. Rubí, B.; Morcego, B.; Pérez, R. Deep reinforcement learning for quadrotor path following with adaptive velocity. Auton. Robot. 2021, 45, 119-134. [CrossRef]

24. Zhang, Y.; Zhang, Y.; Yu, Z. Path Following Control for UAV Using Deep Reinforcement Learning Approach. Guid. Navig. Control 2021, 1, 18. [CrossRef]

25. Wen, H.; Li, H.; Wang, Z.; Hou, X.; He, K. Application of DDPG-based Collision Avoidance Algorithm in Air Traffic Control. In Proceedings of the 2019 12th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 14-15 December 2019; pp. 130-133.

26. Hu, J.; Yang, X.; Wang, W.; Wei, P.; Ying, L.; Liu, Y. Obstacle Avoidance for UAS in Continuous Action Space Using Deep Reinforcement Learning. IEEE Access 2022, 10, 90623-90634. [CrossRef]

27. Zhang, H.; Zhang, J.; Zhong, G.; Liu, H.; Liu, W. Multivariate Combined Collision Detection for Multi-Unmanned Aircraft Systems. IEEE Access 2022, 10, 103827-103839. [CrossRef]

28. Qiu, C.; Hu, Y.; Chen, Y.; Zeng, B. Deep Deterministic Policy Gradient (DDPG)-Based Energy Harvesting Wireless Communications. IEEE Internet Things J. 2019, 6, 8577-8588. [CrossRef]

29. Bagdi, Z.; Csámer, L.; Bakó, G. The green light for air transport: Sustainable aviation at present. Cogn. Sustain. 2023, 2. [CrossRef]

30. Guo, Y.; Liu, X.; Jiang, W.; Zhang, W. Collision-Free 4D Dynamic Path Planning for Multiple UAVs Based on Dynamic Priority RRT* and Artificial Potential Field. Drones 2023, 7, 180. [CrossRef]

31. Sun, J.; Tang, J.; Lao, S. Collision Avoidance for Cooperative UAVs with Optimized Artificial Potential Field Algorithm. IEEE Access 2017, 5, 18382-18390. [CrossRef]

32. Song, J.; Hu, Y.; Su, J.; Zhao, M.; Ai, S. Fractional-Order Linear Active Disturbance Rejection Control Design and Optimization Based Improved Sparrow Search Algorithm for Quadrotor UAV with System Uncertainties and External Disturbance. Drones 2022, 6, 229. [CrossRef]

33. Bauer, P.; Ritzinger, G.; Soumelidis, A.; Bokor, J. LQ servo control design with Kalman filter for a quadrotor UAV. Period. Polytech. Transp. Eng. 2008, 36, 9-14. [CrossRef]

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.

免责声明/出版商注释: 所有出版物中的声明、观点和数据仅代表个别作者和/或贡献者的观点，而不是 MDPI 和/或编辑的观点。MDPI 和/或编辑不承担因内容中提及的任何想法、方法、指导或产品导致的人身或财产损害的任何责任。