线下： 口头报告 ☐ / 海报 ☐ / 相同 ☐**线上：** Zoom ☒ / 预录制视频 ☐

话题： 无人机与图像后处理，用于评估地球'表层过程和地貌

SkySim：基于ROS2的仿真环境：使用大语言模型对无人机集群进行自然语言控制

Aditya Shibu[1], Marah Saleh[2], Mohamed Al-Musleh[3], 和 Nidhal Abdulaziz[4][1]数学与计算机科学学院，[2,3]纺织与设计学院（AIM BeyonD研究实验室），[4]工程与物理科学学院，赫瑞-瓦特大学迪拜校区

电子邮箱：[1]as2397@hw.ac.uk，[2]Marah.Saleh@hw.ac.uk，[3]M.Al-Musleh@hw.ac.uk，[4]Nidhal.Abdulaziz@hw.ac.uk。

摘要：无人机群体在物流、农业和监视领域具有广泛的应用，但要安全、可行地控制它们，需要专业知识。传统静态方法限制了适应性，而大语言模型（LLM）虽能实现自然语言控制，却因缺乏物理基础而生成不安全的轨迹。本文介绍了 SkySim，这是一个基于 ROS2、运行于 Gazebo 的仿真框架，可将大语言模型的高层规划与低层安全保障解耦。SkySim 使用 Gemini 3.5 Pro，将用户指令（例如"形成一个圆形"）转化为空间航点，并参考实时无人机状态。人工势场（APF）安全过滤器仅进行最小幅度的调整，以实现碰撞规避、满足运动学限制并遵守地理围栏，确保以 20 赫兹的频率可行地执行指令。

通过对由 3 架、10 架和 30 架 Crazyflie无人机组成的群体进行实验，验证了空间推理准确率（在测试的所有几何基元上均达到 100%）、实时碰撞预防能力和可扩展性。SkySim 让非专业人员能够迭代优化行为，为动态环境中的 AI 认知与机器人安全搭建桥梁。未来工作将着眼于硬件集成。

关键词：微型与小型无人机控制、大语言模型（LLM）、无人机集群、自然语言控制、ROS2、Gazebo 仿真

1. 引言

无人机集群是功能强大的工具，能够执行涵盖物流、农业和监视等多个领域的任务 [1]。然而，设计复杂的动作或控制此类系统仍是一项艰巨挑战，需要丰富的专业知识，才能确保安全性与物理可行性之间的平衡 [2]。过去，无人机系统一直依赖静态控制方法，例如预编程任务流程和基于规则的库，这使其只能执行预定义的动作，并限制了其在动态真实世界环境中的适应性[3]。大语言模型（LLM）的兴起带来了潜在的解决方案，使用户能够通过自然语言将高层次的人类意图转化为可执行的控制指令 [1][4]。然而，大语言模型生成的参考轨迹本质上并不安全，因为大语言模型缺乏必要的现实世界落地能力，无法理解物理学或预测碰撞等事件；尤其是在无人机集群中，即使推理出现微小错误，也可能立即导致任务失败 [2][4]。

为了在维护物理安全的同时实现自然语言控制的优势，需要采用解耦机制来强制执行安全约束。在本文介绍了**SkySim**，这是一个基于 ROS2 的仿真框架，通过将大语言模型的高层规划与低层实时安全滤波器解耦，弥合了这一差距。该系统利用大语言模型的推理能力控制无人机集群，同时在其底层运行基于优化的安全滤波器；当轨迹违反约束时，该滤波器仅进行最小幅度的调整，从而确保轨迹可行、无碰撞，并适用于真实世界部署。该架构使非专业人员能够使用自然语言迭代改进集群行为，而无需承担处理安全性或低层可行性约束的负担。

1.1. 相关工作

将大语言模型集成到无人机系统中，已成为弥合高层人类意图与可操作控制指令之间差距的变革性方法。近期文献表明，大语言模型使用户能够发出自然语言指令，并将其转换为可执行轨迹，从而超越传统静态框架的局限性 [1-3]。然而，大语言模型生成结果所面临的可靠性挑战以及缺乏现实世界落地，要求开发

混合架构，以确保任务关键型应用中的安全性[4-6]。

该领域的一个领先范式是 SwarmGPT 框架[7]，，它率先将大语言模型集成用于同步的无人机群编舞。SwarmGPT 通过为大语言模型增配基于优化的安全滤波器来应对根本性的安全挑战，该滤波器以最小幅度修正所生成的轨迹，以确保其保持无碰撞。然而，该架构主要针对时间协调编队编排进行优化，侧重于群体运动随时间推移的平滑性与美学品质。它对零样本空间推理的关注较少；在这一任务中，主要目标是在未针对这些特定形状进行预先训练的情况下，将抽象的自然语言指令快速转化为有效的任意静态几何形状（例如，“组成一棵圣诞树”）。因此，聚焦实时可达集生成与空间优化的替代策略已成为关键补充，凸显了对能够在快速的目标导向重构过程中确保安全性的框架的需求[8], [9]。

此外，安全过滤器的有效性在很大程度上取决于所选的避碰算法 [10]。尽管近期研究凸显了卷积神经网络（CNN）和深度强化学习（DRL）等技术的有效性 [11-14], ，但这些方法往往会显著增加计算开销和训练复杂度。在高层规划器（大语言模型）本身已面临推理延迟波动的架构中，加入计算密集型的基于学习的安全层会增加控制回路延迟的风险，在快速且密集的机动过程中甚至可能导致碰撞。相比之下，人工势场（APF）等确定性反应模型提供了一种轻量且截然不同的替代方案 [11]。利用可通过真实值或动作捕捉系统轻松获取的实时状态信息，人工势场能够以极低的计算成本实现即时、反射式的碰撞规避。这印证了 SkySim 致力于实现的解耦式“感知—规划—执行”架构：由大语言模型负责高层认知推理，同时由快速、确定性的人工势场层实时执行物理安全约束。

2. 系统架构与方法论

我们的方法论基于在机器人操作系统（ROS2 Jazzy）中实现的模块化解耦架构。为确保系统在网络条件变化时仍保持稳定性，该框架严格将高延迟的认知规划与低延迟的实时执行分离。

图 1 展示了这一多层架构，详细说明了规划层（由大语言模型驱动的目标生成）与控制层（以 20Hz 执行安全保障）之间的异步数据流。

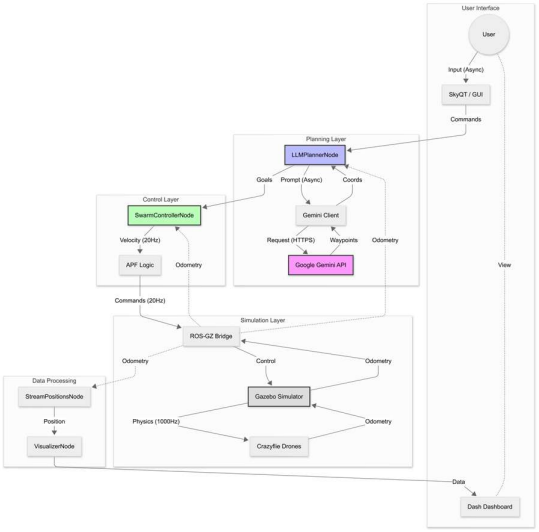


图 1。 SkySim 系统架构。

2.1. 指令输入与状态感知

该流程由一条自然语言命令启动（例如：“在中心周围 2m 高度处形成一个圆形”）。收到输入后，SkySim 接口通过 ROS2 的 /odom 话题订阅所有智能体当前的空间坐标，从而启动“感知”阶段。

(/odom) 话题从 Gazebo 仿真直接桥接至控制节点。此举旨在模拟高精度的真实世界动作捕捉系统（例如 OptiTrack）。这提供了当前群体拓扑的无噪声即时快照。

2.2. 状态感知的基于大语言模型的规划

与标准的大语言模型控制流程不同，SkySim 不会尝试将文本直接映射为离散控制原语（例如油门、俯仰）。相反，它通过 Gemini 3.5 Pro 大语言模型（google.generativeai API）充当高层空间规划器。

为确保输出具有确定性且符合物理可行性，大语言模型会通过严格的系统指令进行初始化，该指令定义了群体规模 *(N)*, 坐标系 *(X*=向前，*Y*=向左，*Z*=向上*)*, 以及地面碰撞约束 *(Z ≥ 0.5 m)*。

当用户发出指令时，框架会将系统指令、所有无人机即时的三维位置遥测数据以及自然语言输入拼接起来，构建动态的状态感知提示。

示例提示结构（N = 3）：系统指令：

“你是一名无人机群体控制器

适用于 3 架无人机。生成目标 [x, y, z] 坐标以执行该命令。仅输出有效的、包含 3 个列表的 Python 列表。保持 Z ≥ *0.5*”。

当前上下文：当前无人机 P

置:[[0.1, 0.0, 1.0], [0.0, 1.5, 1.1], [1.2, 1.1, 0.9]].**用户命令：**

“围绕中心组成一个三角形。”

大语言模型必须输出一个序列化配置矩阵，并按照公式 1 所示的原始 Python 列表格式进行排版：

$$P_{goal} = [[x_1, y_1, z_1], \ldots, [x_N, y_N, z_N]] \quad (1)$$

此字符串通过抽象语法树进行确定性解析，以供验证。不允许添加任何格式设置、Markdown 或文本说明。如果发生 API 超时或输出模式格式错误，系统将保持上一条命令指定的位置，以防止系统状态未定义。

2.2. 安全过滤器与安全执行

“执行”阶段由安全过滤器管理，弥合静态大语言模型航路点与实时动态可行性之间的差距。我们使用单积分器运动学模型对多智能体系统进行建模，其中智能体 i 的控制输入是其速度向量，$u_i = p_i$。

为了在实时环境中实现碰撞规避，同时避免非线性求解器（例如模型预测控制）带来的计算开销，SkySim 实现了一个人工势场控制器。控制回路以 20 赫兹运行，持续将所需速度计算为复合势函数的梯度：

$$v_{total,i} = v_{att,i} + \sum_{j \neq i} v_{rep,ij} \quad (2)$$

1. 吸引速度（目标寻求）： A pr比例控制器被用来将智能体拉向其分配的航点 $P_{goal,i}$。增益为 $K_p = 1.0$：

$$v_{att,i} = K_p \cdot (P_{goal,i} - P_i) \quad (3)$$

2. 排斥速度（碰撞规避）： 系统强制施加安全约束，以防止智能体之间发生碰撞。如果智能体 i 与邻居 j 之间的欧几里得距离 $d_{ij} = \| P_i - P_j \|$ 小于安全半径 $r_{min} = 0.8$ 米（可配置），则会激活线性排斥向量：

$$v_{(rep,ij)} = \begin{cases} K_{rep} \cdot (r_{min} - d_{ij}) \cdot \hat{n}_{ij} & if\ d_{ij} < r_{min} \\ 0 & otherwise \end{cases} \quad (4)$$

其中$n_{ij}\hat{}= \frac{P_i - P_j}{d_{ij}}$ 是指向远离邻居方向的单位向量，$K_{rep} = 2.0$是排斥增益。

3. 运动学与边界约束：为确保绝对的系统安全，指令在执行前会通过两阶段约束协议进行净化。首先，LLM规划器节点强制执行严格的虚拟地理围栏。任何由大语言模型生成且落在既定空间边界($X, Y \in [10.0, 10.0]\ m, Z \in [0.2, 5.0]\ m$)之外的航点都会被立即拒绝，并触发位置保持回退机制，以防止群体离开动作捕捉系统的仿真范围。

其次，对于有效工作空间内的航点，会施加纳米四旋翼无人机的运动学限制。最终生成的速度指令在发布到 ROS2（/cmd_vel）话题之前，会被限制在 $v_{max} = 0.5$ m/s：

$$v_{cmd,i} = \min(|v_{total,i}|, v_{max}) \cdot \frac{v_{total,i}}{|v_{total,i}|} \quad (5)$$

这种多层架构可确保群体始终被限制在安全的运行空间内，并始终遵守电机物理限制。

3. 实验评估与结果

为验证 SkySim 框架的安全性、可扩展性和认知能力，我们使用 Gazebo Harmonic环境和 ROS2 Jazzy 开展了一系列仿真实验。群体由模拟的 Crazyflie 2.1无刷版无人机组成。无人机的位置直接从 Gazebo 提供的真实值中获取，数据通过 ROS2（/odom）话题发布。

评估重点关注三项主要指标：大语言模型空间推理准确率、实时碰撞规避和系统可扩展性。我们测试了群体规模为（N = 3、10 和 30）的智能体，以评估该框架在密度增加时的表现。

3.1. 测试场景与群体可扩展性

为评估该框架，我们设计了三个不同的实验场景，用于测试大语言模型的空间推理、实时安全性和边界约束。系统在（N = 3、10 和 30）种群体规模下动态扩展，以评估人工势场控制器的计算能力限制、LLM上下文窗口以及该框架的可扩展性。

场景A：几何队形控制（空间推理）该群体被指挥执行高层次的

拓扑任务，包括：“形成一个半径为2米的圆形。”，“形成一个半径为2米的球体。”，“在高度为2米处形成一个5×6的网格。”，“形成一个三维立方体。”，以及“形成一棵圣诞树。”。

Gemini 3.5 Pro Preview 成功将这些抽象指令转换为有效的*N*维空间坐标。在所有群体规模下（N = 3 至 30），大语言模型在定义的几何基元上取得了 100% 的成功率，并为每项测试准确计算出目标坐标。系统展现出强大的可扩展性：大语言模型在 ≈ 50 秒内生成可执行计划（对于 N=30 ），人工势场控制器则成功引导智能体收敛至目标队形，且未发生碰撞。

图 2. 通过大语言模型提示生成的无人机队形：（左）N = 3 个无人机组成的三角形，（中）N = 10 个无人机组成的圆形，半径为 3 m，（右）N = 30 5×6 个无人机组成的网格。提示：

- ***N* = 3：**“在高度 2m 处，以零点为中心组成一个三角形。• ***N* = 10：**“在高度 2m 处，以零点为中心组成一个半径为 3m 的圆形。”• ***N* = 30：**“在高度 2m 处组成一个 5x6 网格。”

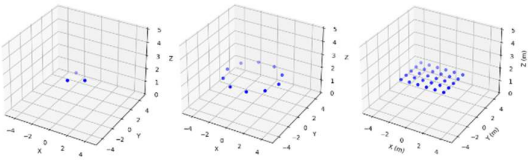


图3。无人机到达大语言模型设定的航路点后的最终位置

场景B：静态与动态避碰（人工势场压力测试）为评估解耦的

人工势场控制器，我们在两项不同的避碰测试中测试了中等群体规模（N= 10）。

静态危险（不安全航路点测试）：我们启动了一项自动化压力测试，命令整个群体汇聚到同一个点。图3展示了智能体间距。如图所示，欧几里得距离仅在很短时间内略微低于0.8米的阈值。这一瞬间触及安全边界，正是触发人工势场控制器的机制；由此产生的排斥向量立即压过目标吸引力。这表明，人工势场控制器成功检测到无人机正在汇聚，并且斥力压过了吸引力。

动态危险（位置交换测试）：群体收到了一条自然语言命令：“所有无人机与正对着自己的无人机交换位置。”这条命令迫使所有智能体轨迹同时在工作空间中心动态交汇。如果没有底层安全过滤器，这将导致灾难性的多智能体碰撞。然而，斥力场（r_{min} = 0.8）实时动态地引导智能体相互绕行，使群体能够化解冲突并安全抵达目标航路点。

3.2 安全性与可行性指标

为了严格量化安全过滤器的性能，我们在整个仿真运行时间内监测了两项关键指标：用于碰撞验证的最小智能体间距离（d_{min}），以及所有不同智能体对的群体速度曲线。我们专门分析了高密度结构形成期间的这一指标：N = 10 “圣诞树”、N =30“圣诞树”（复杂三维拓扑），以及 N = 30 “星形”（高密度平面拓扑）。

图3展示了最终的编队快照及其距离曲线。需要注意的是，0.8m处的红色虚线表示APF激活阈值，而非物理碰撞边界。如图所示，为稳定密集编队，d_{min} 自然会降至该保守缓冲区以下，但始终严格高于物理碰撞极限（r_{drone} ≈ 55mm [15]），从而验证了实时安全性。

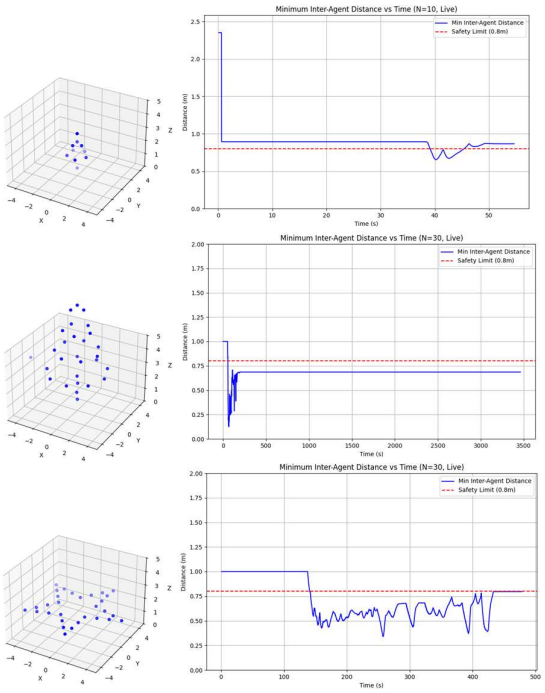


图3.（左）三维群体快照。（右）随时间变化的智能体间最小距离。红线（0.8m）表示APF激活缓冲区，未发生物理碰撞。

为确保生成的轨迹平滑且可飞行，我们分析了 N=10 群体在重构过程中的速度曲线，如图 4 所示。该曲线显示，加速阶段平滑，随后自然减速，没有高频振荡。峰值速度（≈ 0.42 m/s）始终严格处于规定的运动学限制（v_{max} = 0.5 m/s）之内，验证了解耦架构能够生成物理上可行的控制指令。

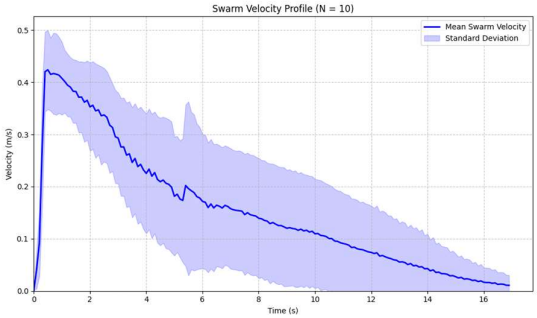


图 4. N = 10 配置下无人机的速度曲线

3.3. 延迟分析

为评估该框架在大规模作业中的计算可行性，我们测量了端到端规划延迟，其定义为从发出自然语言命令到收到大语言模型解析后的坐标列表所经过的总时间。

图 5 展示了不同群体规模下的延迟分布（N = 3、4、10、30）

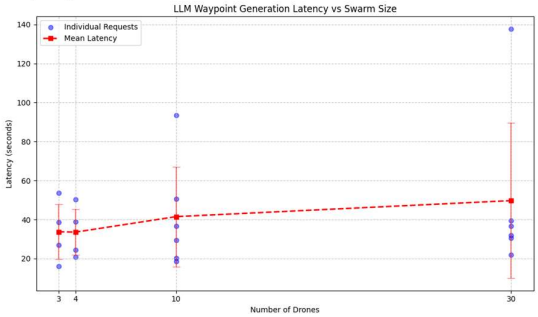


图 5. 端到端规划延迟与群体规模的关系。

结果凸显了两个关键性能特征：

次线性可扩展性： 系统展现出良好的集群密度可扩展性。尽管群体规模扩大了 10 倍（N = 3 至N =10），平均规划延迟仅增加了约 45%（34 秒增至 50 秒）。这表明处理无人机状态数据的计算成本能够高效增长，证明系统可以处理更大规模的群体，而不会导致处理时间大幅增加。

解耦安全： 规划时间存在显著差异（例如，N = 30 时出现的离群值），这是受 API 服务器负载和网络抖动影响的基于云的概率模型的典型特征。然而，规划延迟（秒）与控制回路延迟（毫秒）完全解耦。如图 1 所示，本地安全过滤器以 20 Hz 持续运行，确保这些不定的规划延迟在规划阶段不会危及群体的物理安全或稳定性。

4. 讨论与局限性

实验结果验证了 SkySim 将自然语言转化为安全、可行的集群行为的能力。具体而言，在测试的几何基元上达到 100% 的成功率，证明了大语言模型的空间推理能力，在适应性和易用性方面也有效优于静态控制库。

影响： 事实证明，集成解耦安全过滤器至关重要。尽管大语言模型生成了有效的目标状态，但智能体间距离短暂下降，凸显了人工势场的反应式特性。该机制在所有N= 3、N = 10和N = 30 次试验中都有效防止了碰撞，但代价是在高密度机动期间收敛时间增加。通过将高层规划与低层安全解耦，SkySim 降低了安全关键领域的风险，有望推动无人机应用于农业或灾害响应等领域，使非专业人员无需掌握大量编程知识，也能按需实现复杂的集群行为。

局限性： 尽管取得了这些成果，当前架构仍存在若干亟待解决的局限。系统依赖基于云的推理（Google Gemini API）。如延迟分析（图 5）所示，这会导致显著的延迟波动（异常值 > 100s），其主要原因是服务器负载。虽然本地安全机制可在这些延迟期间保障物理安全，但目前该系统并不适用于需要在毫秒级时间内响应新指令的时间关键型任务。

SkySim 还依赖 Gazebo 提供无噪声的真值状态估计。真实世界部署会引入传感器噪声（例如来自 Optitrack [16] 或 GPS 的噪声）、通信丢包和环境扰动（例如天气状况），这些因素可能会降低人工势场排斥向量的精度。

尽管系统的扩展性呈次线性，但大语言模型的上下文窗口限制了群体规模上限。对于N > 30，编码每个智能体位置所需的提示长度可能会超过词元限制；由于需要足够的本地硬件资源，以维持物理引擎的实时保真度以及所有智能体之间的 ROS2 节点通信，我们的验证规模上限为 N = 30 。

确定性人工势场控制器虽然速度快，但容易陷入局部极小值。在高度杂乱的环境或复杂障碍物场中，智能体可能在到达目标之前被困于状态平衡点，这是反应式势场相较于预测求解器的一个已知局限。

最后，该系统针对零样本静态编队进行了优化。它原生不支持连续的时限轨迹（例如“旋转球体”），因此目前仅适用于基于航点的重构。

5. 结论与未来工作

本文介绍了 SkySim——一个基于 ROS2 的仿真环境，旨在弥合自然语言与基于大语言模型的认知之间的鸿沟。通过将概率性大语言模型推理与确定性的 20Hz 安全过滤器隔离开来，我们证明了非专业人员能够指挥复杂的集群编队，同时不损害物理安全。

未来工作将聚焦于三个关键领域，以解决已发现的局限性

1. 硬件集成：我们计划将整套系统移植到实体 Crazyflie 2.1 无刷无人机上[15]，使用 OptiTrack 动作捕捉系统 [16] ，以量化真实环境中传感器噪声的影响。

2. 高级控制求解器：为克服人工势场的局部极小值局限性，我们将探索集成模型预测控制（MPC），将其作为主动式安全层，使其能够比简单的人工势场更有效地处理动态约束。

3. 多模态输入：未来版本将加入多模态能力，使群体能够响应视觉输入或动态环境变化，进一步将“感知—规划—行动”循环扩展到静态坐标之外。

4. 边缘部署与优化：为消除云端推理固有的延迟瓶颈，未来版本将从通用型云端大语言模型转向直接部署在本地硬件上的小语言模型（SLM）。我们计划利用高效的微调框架，例如 Unsloth [17]，，针对集群控制词汇对这些模型进行专门优化。此举旨在将端到端规划时间缩短至亚秒级，使系统能够结合模型预测控制，支持目前因基于 API 的延迟而无法实现的动态高速机动。

参考文献

[1]. G. Aikins、M. Dao、J. Moukpe、T. Eskridge 和 K. Nguyen，《LEVIOSA：基于自然语言的无人航空器轨迹生成》，《Electronics》，第 13 卷，第 22 期，第 4508 页，2024年。[2]. Y. Ping 等人，《多模态大语言模型赋能的无人机集群：迈向高效智能的自主空中系统》，arXiv 预印本 arXiv:2506.12710，2025年。[3]. B. Han、Y. Chen、J. Li、J. Li 和 J. Su，《SwarmChain：面向无人机集群控制的协同大语言模型推理》，《IEEE物联网杂志》，第 8 卷，第 5 期，第 64–71 页，2025年9月。[4]. L. Yuan、C. Deng、D. Han、I. Hwang、S. Brunswicker 和 C. Brinton，《面向无人机的新一代大语言模型：从自然语言到自主飞行》，arXiv 预印本 arXiv:2510.21739，2025年。[5]. S. Sarkar 等人，《基于大语言模型的控制综述》，2025年，doi：10.13140/RG.2.2.28595.54567。[6]. H. Kurunathan、H. Huang、K. Li、W. Ni 和 E. Hossain，《机器学习辅助的无人机运行与通信：当代综述》，《IEEE Communications Surveys & Tutorials》，第 26 卷，第 1 期，第 496–533 页，2024年。[7]. M. Schuck 等人，《SwarmGPT：将大语言模型与安全运动规划相结合，实现无人机群编舞》，arXiv 预印本 arXiv:2412.08428，2025年。[8]. P. Xie、S. Diaconescu、F. Stoican 和 A. Alanwar，《环岛约束凸生成器：多连通可达集的统一框架》，arXiv 预印本 arXiv:2511.07330，2025年。[9]. Z. Wang 和 H. Ding，《用于移动机器人路径规划的反向学习平衡优化器》，《International Journal of Robotics and Automation Technology》，第 10 卷，第 64–74 页，2025年。[10]. P. Mishra、B. Boopal 和 N. Mishra，《利用机器学习实现无人机实时三维航路优化》，《ICST Transactions on Scalable Information Systems》，2024年。[11]. J. Yasin、S. Mohamed、M. Haghbayan、J. Heikkonen、H. Tenhunen 和 J. Plosila，《无人机：避碰系统与方法》，《IEEE Access》，第 8 卷，第 105139–105155 页，2020年。[12]. X. Dai、Y. Mao、T. Huang、N. Qin、D. Huang 和 Y. Li，《基于卷积神经网络学习的四旋翼无人机自动避障》，《Neurocomputing》，第 402 卷，第 346–358 页，2020年。[13]. R. Bălaşa、M. C. Bîlu 和 C. Iordache，《基于近端策略优化的无人机姿态控制强化学习方法》，《Land Forces Academy Review》，第 27 卷，第 4 期，第 400–410 页，2022年。[14]. J. Hu、L. Wang、T. Hu、C. Guo 和 Y. Wang，《基于深度强化学习的双无人机协同空战自主机动决策》，《Electronics》，第 11 卷，第 3 期，第 467 页，2022年。[15]. Bitcraze AB，《Crazyflie 2.1无刷版数据手册》，瑞典斯科讷省马尔默。[在线]。可访问：https://www.bitcraze.io/documentation/hardware/crazyflie_2_1_brushless/crazyflie_2_1_brushless-datasheet.pdf。[访问日期：1月。24, 2026].[16]. NaturalPoint公司，《PrimeX 22 技术规格》，美国俄勒冈州科瓦利斯。 [在线]。可访问：https://optitrack.com/cameras/primex-22。[访问日期：1月。24, 2026].[17]. Unsloth AI，《Unsloth 文档》， [在线]。可访问：https://unsloth.ai/docs。[访问日期：1月 24日，2026]。