arXiv:2302.05849v1 [cs.MA] 12 Feb 2023

城市空中交通垂直起降机场多航空器起降的基于图学习的决策支持

Prajit KrisshnaKumar[1*]、Jhoel Witter[1*]、Steve Paul[1],Karthick Dantu[2],Souma Chowdhury[1]

摘要——**预计城市空中交通（UAM）概念下的大多数航空器将采用电动垂直起降飞行器（eVTOL），并在垂直起降机场运行。这类似于通用航空航空器与机场之间的关系，但垂直起降机场规划建于人口稠密的城市环境中，因此管理其服务范围内的空中交通面临独特挑战。随着定期降落和起飞频率不断增加，这一挑战将愈发突出。本文假设由中央空中交通管制员（ATC）进行管制，以探究一种新型人工智能驱动的空中交通管制方法在管理垂直起降机场所服务的电动垂直起降飞行器方面的表现。由最小间隔要求驱动的安全性与延误是此场景下的两个重要考量因素。本文将空中交通管制问题建模为任务分配问题，并将通信中断（例如链路质量差）和恶劣天气（例如强阵风效应）造成的不确定性设定为动作失败的小概率事件。为学习垂直起降机场的空中交通管制策略，本文提出了一种名为"城市空中交通—垂直起降机场调度管理（UAM-VSM）"的新型基于图的强化学习（RL）方案。该方法利用图卷积网络（GCN）将垂直起降机场空间和电动垂直起降飞行器空间抽象为图，并汇总信息，以帮助中央空中交通管制智能体对环境进行泛化。仿真环境采用 Unreal Engine 与 AirSim；训练和测试均在其中进行。由于在此类高真实度仿真中进行蒙特卡洛采样的成本很高，不确定性仅在测试期间纳入考虑。与可行的随机决策基线和先到先服务（FCFS）基线相比，所提出的基于图的强化学习方法在测试场景中表现显著更佳，同时还能泛化至未见场景，并适应不确定性。**

I. 引言

交通运输技术日新月异，自动驾驶汽车和自主空中包裹配送也即将成为现实。据估计，到 2050 年，全球约 68% 的人口将居住在城市区域 [1]。城市空中交通（UAM）为交通方式增添了新的维度，即利用垂直起降（VTOL）设备在中等高度运送人员[2]。城市空中交通的概念可追溯到 1953 年，当时纽约航空公司运营直升机空中出租车。随着电力、推进系统和电池领域的最新进展，空中出租车正变得更具可行性且更经济 [3]。Uber 等公司正竞相在城市区域开发和部署垂直起降飞行器 [4]，但由于部署面临包括政府监管在内的多项挑战，具体时间仍未可知。鉴于现有交通空间的可用性以及预估的人口规模，城市空中交通在不久的将来势在必行。

目前，空中交通管制（ATC）负责管理所有具有飞行能力的飞行器 [5]。垂直起降机场是垂直起降飞行器起飞、降落并为电池充电的区域 [6]。对于垂直起降飞行器而言，每小时进出垂直起降机场的飞行器数量可达数百至数千架 [7]。在这种情况下，管理航空器的降落/起飞（L/TO）更具挑战性，也引发了对安全性和监管的担忧。在存在不确定性和紧急情况时，先到先得的理念并不适用。本文提出了一种方案，用于管理垂直起降机场区域内的垂直起降飞行器，同时保障其安全性。这是一个多维问题，需要采用超越简单线性数学建模的方法来建模。因此，我们创建了一种新颖的三维模拟环境，其中融入了真实物理特性以及路径规划等基本要素。值得注意的是，我们的模拟器运行速度远快于实时，这有助于我们收集学习所需的数据。我们训练了两种不同的基于学习的算法，并对其进行了比较。已有多项研究探讨了垂直起降机场 [8]和垂直起降飞行器 [9]的建模。垂直起降机场和垂直起降飞行器的设计不在本文讨论范围内；相反，我们将停机位视为直升机停机坪，将无人机视为垂直起降飞行器，因为这一概念同样适用。

近年来，基于人工神经网络（ANN）的学习算法被用于各种智能自主系统中，与人类相比，它们在决策方面发挥着更好的作用 [10]。许多成功的强化学习应用，例如自动驾驶汽车和机器人技术，都涉及多个智能体，并被作为多智能体强化学习（MARL）问题求解 [11]。过去几年中，已有多项值得关注的研究将基于图的强化学习（RL）应用于各种单智能体和多智能体组合优化（CO）问题 [12]–[24]。在这些研究中，可建模为图的状态空间变量由图神经网络（GNN）编码，并作为策略网络的一部分。多智能体强化学习面临若干难题，例如

*作者贡献相同[1] 布法罗大学机械与航空航天工程系，美国纽约州布法罗市，邮编 14260。[2]布法罗大学计算机科学与工程系，美国纽约州布法罗市，邮编 14260。

† 通讯作者，soumacho@buffalo.edu本研究由 Stephen Still 可持续交通与物流研究所（SSISTL）资助。本文所表达的任何意见、研究结果、结论或建议均为作者观点，不一定反映 SSISTL 的立场。

维数灾难或状态-动作空间的指数级增长、非平稳性——复杂的动态变化，以及信用分配问题——即难以确定应奖励哪个智能体 [25]。近期一些克服这些局限的研究，将多智能体强化学习问题转化为单智能体问题 [26] （集中式训练）：收集所有智能体的经验，由一个智能体进行训练；随后采用分布式实施，将训练好的模型应用于所有智能体，以实现分布式决策 [27]。我们将问题建模为单智能体问题，不过智能体数量可以超过本文所涉及的数量。此外，我们将连续环境离散化，并以离散空间构建状态空间；与连续状态空间相比，这样更易于学习。

本文的主要贡献包括：1）将垂直起降机场运营管理建模为马尔可夫决策过程（MDP），即短期尺度的垂直起降飞行器降落/起飞（L/TO）调度问题；2）开发用于模拟城市空中交通垂直起降机场运行的三维仿真环境；3）开发一种（图）学习框架，以制定策略，在考虑环境和运行不确定性的情况下，确定航空器在垂直起降机场运行空间内起降的时机。本文其余部分安排如下：第 II 节介绍 MDP 建模和学习方法。第 IV 节简要介绍仿真架构、各组成部分的运行方式，以及所采用学习算法的细节。第 V-B 节介绍不同的案例研究。

II. 垂直起降机场运营管理：建模与学习方法

这是一个城市空中交通垂直起降机场时间表管理问题，目标是设计一种 GCN 策略，用于训练空中交通管制智能体。该智能体必须能够：为其空域中的电动垂直起降飞行器分配任务（充电、起飞、降落、悬停），维持所有电动垂直起降飞行器的高电量水平，避免碰撞，并遵循每架电动垂直起降飞行器指定的飞行计划。环境由 2 个普通停机位、1 个充电停机位、智能体空域之外的 5 个目的地以及 7 个悬停点组成。图 1 展示了该环境的简化移动示意图。接下来的小节将详细介绍该环境以及我们为其选择的马尔可夫决策过程（MDP）建模
这个问题。

A. 环境

环境中初始化了 4 架电动垂直起降飞行器，每架都会起飞前往随机目的地，然后返回。每架电动垂直起降飞行器降落在垂直起降机场后，都会收到新的飞行计划，其中指定了其需要起飞的时间（即未来 10 到 20 分钟内）以及需要前往的目的地。电动垂直起降飞行器抵达目的地后，会自动收到返回垂直起降机场的飞行计划，而空中交通管制需要在其抵达后的 15 分钟内安排其降落。

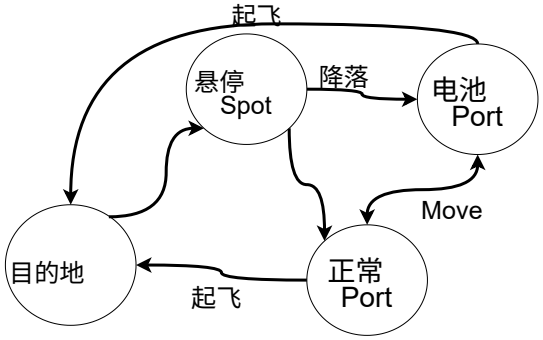

图 1：垂直起降机场环境的状态与动作示意图

图 4 对此作了进一步说明。每架电动垂直起降飞行器起始时电量充足，并会在每个时间步消耗电量。放电速率取决于电动垂直起降飞行器的状态：

$$\text{discharge rate} = \begin{cases} \text{distance traveled} \times 0.5 & \text{if cruising} \\ 2 & \text{if hovering} \\ 4 & \text{if idling on ground} \end{cases} \tag{1}$$

这样做是为了确保电动垂直起降飞行器在静止时会消耗电池电量。每一步中，如果电动垂直起降飞行器停在充电停机位上，就可以恢复 10% 的电池电量。每架电动垂直起降飞行器异步移动，因此可能有多架电动垂直起降飞行器同时移动。通过这种配置，智能体可以学会避免两架或更多电动垂直起降飞行器相交时发生碰撞。每个时间步，智能体都会按顺序选择一架新的电动垂直起降飞行器执行动作。如果选中的电动垂直起降飞行器当前位于目的地，智能体会等待它重新进入垂直起降机场空域，然后再为其选择动作。模拟以实时速度的 300 倍运行，因此模拟器中每经过一秒，就相当于经过 5 分钟。这有利于训练，因为每步耗时接近 0.2 秒，每回合约为 288 秒（4.8 分钟），共 1440 步。模拟中的每个回合约为 24 小时，即完整的一天运营时间。每架电动垂直起降飞行器至少以 $45Hz$ 的频率更新，其中包括更新所有特征（位置、延误、电池状态和电动垂直起降飞行器状态）。

B. 马尔可夫决策过程建模

该问题被建模为一个马尔可夫决策过程（MDP），其参数如下：

$$\begin{aligned} s &= \mathcal{F}_s(b_i, c_i, l_i, x_{i,p}, y_{i,p}, p_{a,t}) \\ a &= \mathcal{F}_a(s) \\ r &= \mathcal{F}_r(a, s, \S, \beta, \gamma, \tau, \gamma, w_n) \end{aligned} \tag{2}$$

其中， i 和 p 分别代表所有飞行器和停机位。状态空间和动作空间见表 I，奖励如公式 3 所示。τ 为起飞系数， γ 为着陆系数， λ 为电池系数， β 为延误系数， $\S$ 为安全系数， w_n 为权重。这些系数将在下文进一步解释。

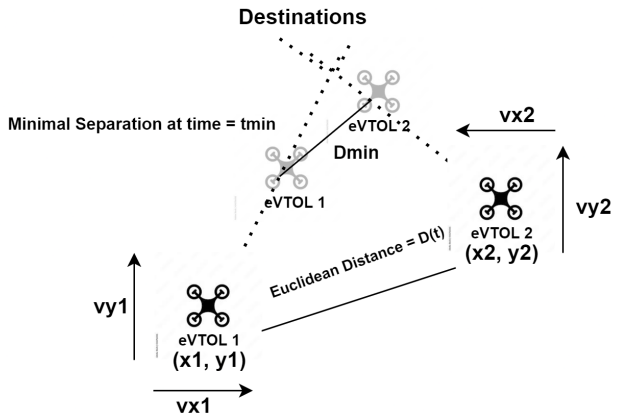

图 2：最小间隔场景

Type	变量
垂直起降机场状态	可用性 - P_a 机场类型 - P_t 位置 - (x_p, y_p)
垂直起降状态	当前状态 - c_i 电池容量 - b_i 调度状态 - l_i 位置 - (x_i, y_i) 端口可用性 - P_a
动作空间	保持静止 起飞 移动/降落至普通停机位 - 1,2 移动/降落至电池端口 - 1 移动至悬停点 - 1,2,3,4,5,6,7 继续先前动作 避免碰撞

表 I：MDP 建模

$$R = w_1\tau + w_2\gamma + w_3\lambda + w_4\beta + w_5\S \tag{3}$$

1) 起飞与降落系数：我们将**“良好”**的起飞定义为电动垂直起降飞行器满足以下条件：**i)** 按时起飞（在其计划起飞时间前后 5 分钟内）；**ii)** 起飞时电池电量高于 30%。对于**“良好”**的降落，标准相同，但电动垂直起降飞行器还可以选择早于其计划时间降落。 τ 和 $\gamma \in \{-5, 5\}$。

2）电池系数：该系数定义如下：

$$\lambda = \begin{cases} 5 \times \frac{battery_remaining}{100} & \text{if } battery_remaining \geq 30 \\ -5 & else \end{cases} \tag{4}$$

该值会随着电动垂直起降飞行器充电而逐渐增加。为进一步避免飞行器在临界电池电量下飞行，当电池百分比低于 30% 时会施加惩罚。要获得电池系数，智能体需要让每架电动垂直起降飞行器都保持满电。

3）延误系数：当电动垂直起降飞行器错过起飞或降落时段后，环境中便会产生延误。随着仿真时间推移，延误会不断增加，直到电动垂直起降飞行器出发并获得新的时间表。延误系数定义如下：

$$\beta = -5 + 10 \times e^{-delay} \tag{5}$$

其中， $delay$ 项的单位为分钟。这促使智能体尽可能减少延误，以获得最大的延误系数。

4) 安全系数：在计算安全系数之前，环境会检查所选电动垂直起降飞行器是否：**i.)** 当前正飞往某个位置；**ii.)** 与另一架正在途中飞行的电动垂直起降飞行器存在相交路径。这可以通过图 2 直观地展示。图中，两架电动垂直起降飞行器正飞向一个交点。它们在任意时刻的距离可表示为时间函数结合它们的瞬时位置和速度向量，使用欧几里得距离计算：

$$D(t) = \sqrt{(x_1 - x_2 + tv_{x_1} - tvx_2)^2 + (y_1 - y_2 + tv_{y_1} - tvy_2)^2} \tag{6}$$

其中， x_n、 y_n、 v_{x_n}、 v_{y_n} 分别是电动垂直起降飞行器 1 和 2 的位置与速度分量。随后对该方程关于时间求导，并求解局部最小值 t_{min}:

$$t_{min} = -\frac{2(v_{x_1} - v_{x_2})(x_1 - x_2) + 2(v_{y_1} - v_{y_2})(y_1 - y_2)}{2(v_{x_1} - v_{x_2})^2 + 2(v_{y_1} - v_{y_2})^2} \tag{7}$$

随后将 t_{min} 代回方程 6，以获得最小间隔 D_{min}。每架模拟电动垂直起降飞行器的占用空间为 1x1 米，因此，如果 D_{min} 小于 3 米且智能体不采取规避动作，该智能体将受到惩罚：

$$\S = \begin{cases} 0 & \text{if } D_{min} \; is \; None \\ -5 & \text{if } D_{min} \leq 3.0 \; \& \; action \neq avoid \; collision \\ 5 & \text{if } D_{min} \leq 3.0 \; \& \; action = avoid \; collision \end{cases} \tag{8}$$

5) 奖励权重：每个系数都会根据其重要性乘以一个权重 w_1、 w_2、 $..w_n$ 。在我们的问题中，安全性被认为更为重要，因此为安全系数 § 分配了最大权重，其次是 β、 τ、 γ、 λ。

III. 学习架构

本文聚焦于一种称为近端策略优化（PPO）的深度强化学习框架 [28]。PPO 与信赖域策略优化（又称 TRPO）类似 [29]。PPO 的独特之处在于，它利用特有的目标函数对策略扩展进行裁剪，从而无需付出策略更新幅度过大且不稳定的代价，便能更安全地进行策略探索。通过增大或减小裁剪参数，可以控制更新幅度；再结合 Adam 梯度下降，便构成了一种强大的学习算法。我们使用 OpenAI Gym 和 Stable Baselines 3 [30] [31] 进行强化学习，并且

训练环境的总体流程见图 5。

*A. GCN*智能体

如上所述，PPO 是一种先进的演员-评论家强化学习方法，已展现出高效率、广泛的适应性和可靠性 [32]。本文将使用图学习PPO智能体，该智能体通过由图神经网络（GNN）构成的策略网络进行训练。过去几年中，GNN 已成功应用于各种任务分配、调度和路径规划问题 [17], [24],[33] 。GNN 的主要优势之一是能够利用以图结构数据表示的问题的结构信息（局部和全局），并将其表示为图嵌入、节点嵌入或边嵌入。在本研究中，我们采用图卷积网络（GCN） [34] 生成图嵌入，并使用定制的多层感知机（MLP）将最终特征向量转换为一组对数概率。用于训练网络的强化学习参数见表 II，GRL智能体使用的网络架构如图 3 所示。图 3 所示的 GRL智能体 MLP 包含两层，分别有 128 个和 64 个神经元。此外，智能体采用掩码，其设置取决于所选电动垂直起降飞行器的状态以及各停机位的可用性。这样可以降低问题的复杂度，使智能体能够专注于其他环境因素，例如避免碰撞和减少不确定性。如图 3 所示，GCN智能体包含特征抽象、策略网络和价值网络，以配合 PPO 使用。我们在线性层中使用偏置，并采用斜率范围为 0.1 至 0.3 的随机化 ReLU（RReLU），因为这是缓解梯度消失最快且最有效的选择。起初我们采用 LeakyReLU，但为每个网络调试激活层非常耗时。反向传播使用学习率为 1e-5 的 Adam优化器。主要区别在于特征抽象网络：我们使用两个图卷积网络，它们以电动垂直起降飞行器和垂直起降机场特征矩阵及其各自的边连接矩阵作为输入。策略网络采用四层 MLP，并通过 log-softmax变换为 11 个动作计算对数概率。智能体采用掩码，其设置取决于所选电动垂直起降飞行器的状态以及各停机位的可用性。这样可以降低问题的复杂度，使智能体能够专注于其他环境因素，例如避免碰撞和减少不确定性。

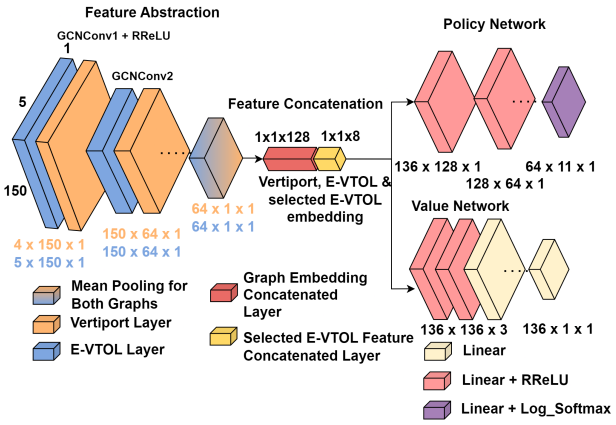

图 3：所提出的 GRL 智能体的特征抽象、价值网络和策略网络

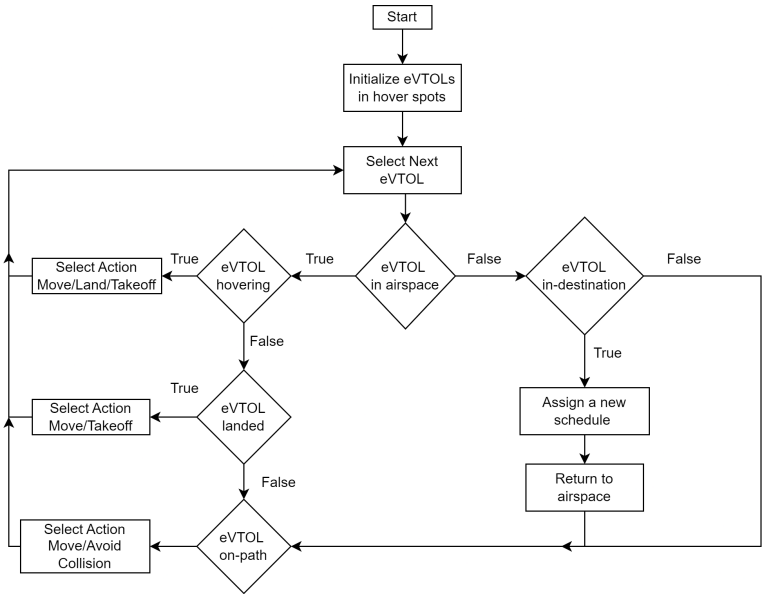

图 4：决策流程图

IV. 仿真环境

我们基于 Microsoft AirSim 开发了一个用于电动垂直起降飞行器的定制仿真系统。AirSim [35]是一个开源机器人仿真平台。AirSim 帮助我们解决训练所需大规模数据集的问题，并支持在仿真器中进行调试。AirSim 利用当前游戏引擎（虚幻引擎）[36] 的渲染、物理特性和感知计算功能，创建准确的现实世界仿真。基于高效生成的真实值数据，这种逼真性使我们能够研究并执行现实世界中复杂、耗时且具有风险的任务。AirSim 使我们能够模拟电动垂直起降飞行器的物理特性，而垂直起降机场和电动垂直起降飞行器的属性则通过 *Python* 以编程方式实现。用于学习的整体框架如图 5 所示。基于 OpenAI Gym 的强化学习接口用于实现 AirSim 与学习智能体之间的通信。下面将对各个组件进行更详细的介绍。

A. OpenAI Gym 接口

Gym 是一个开源的*Python* 库，提供标准 API，用于在学习算法与环境之间进行通信 [37]。自发布以来，Gym API 已成为训练和开发强化学习问题的领域标准。在我们的场景中，Gym 接口在学习框架各组件之间的通信中发挥着关键作用。例如，Gym 接口接收来自动作管理器的解码后动作，并将该信息发送

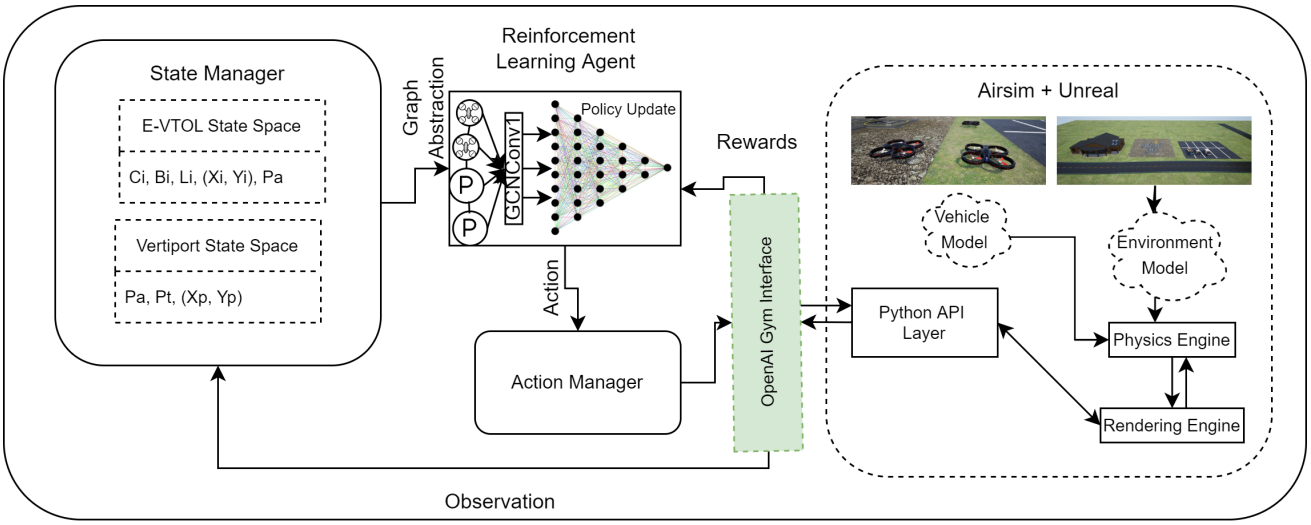

图 5：仿真环境概览，展示信息在状态管理器、动作管理器、学习策略网络和仿真引擎（AirSim + Unreal）之间的流动

到 AirSim 的*Python* API 层，由该层模拟飞行器的物理特性和动力学。

B. AirSim+Unreal

基于虚幻引擎 [36]构建了一个自定义环境，其中包含 3 个垂直起降机场，并集成了配备 5 架垂直起降飞行器的 AirSim 插件。AirSim 插件负责管理垂直起降飞行器的物理特性，并将位置、碰撞信息等数据发送至 OpenAI Gym 接口。AirSim 支持以远快于现实世界时钟的速度运行仿真，这有助于加快训练过程。AirSim API 层接收来自 Gym 接口的指令，例如前往指定位置、起飞或降落；AirSim 负责管理物理特性和路径规划，而虚幻引擎则在屏幕上渲染这些内容。图 6 展示了图形界面。

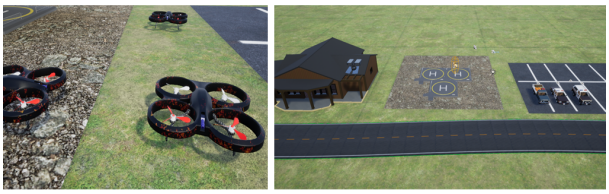

图 6：使用虚幻引擎开发的仿真环境。

C. 状态管理器

状态管理器继承了停机位和垂直起降飞行器的所有属性，负责提取 Gym 接口所请求的特定垂直起降飞行器所需的状态信息。垂直起降飞行器的属性包括：1. 电池电量；2. 时间表；3. 当前位置；4. 状态（准时、延误）。我们定义了两个*Python* 类，一个用于城市空中交通，另一个用于停机位。每架垂直起降飞行器都将继承城市空中交通类的属性，而停机位类则封装垂直起降机场的属性。停机位类负责管理垂直起降机场，并负责向所有飞行器发送各个停机位的状态，并在飞行器降落或离开停机位时更新状态。这里考虑两种停机位：1. 电池接口，只有这类停机位能够为飞行器充电；2. 普通停机位。悬停点也由停机位类管理，飞行器进入停机位区域时会占据其中一个位置，并在此处悬停，直到智能体决定采取某项操作。

D. 动作管理器

动作管理器继承停机位和垂直起降飞行器的所有属性，并负责解码强化学习智能体发送的动作。动作解码后会发送给飞行器，并相应地更新飞行器和停机位的状态。Gym 接口负责在强化学习智能体与动作管理器之间进行通信。

算法	PPO
最大时间步数	300,000
学习率	1e-5
折扣因子	1
步数	20,000
批量大小	10,000
熵系数	0.001
奖励权重	$\{0.3, 0.3, 0.35, 0.1, 0.35\}$

表 II：强化学习参数

V. 结果与讨论

本节分为两个子部分：学习部分和案例研究部分。学习部分将回顾训练情况：哪些方面进展顺利，哪些方面仍需改进。案例研究部分将重点评估 GRL 智能体在环境存在和不存在不确定性时的表现。这里的不确定性包括三方面：

1）在环境中加入风的影响，作为一个对抗向量；该向量会降低电动垂直起降飞行器的线速度，或以 5% 的概率将其完全抵消。2）电池接口在每个时间步有 5% 的概率无法正常工作。3）电动垂直起降飞行器有 5% 的概率无法起飞。

这三项对环境的改动将分别模拟恶劣天气、设备故障以及电动垂直起降飞行器的机械故障。理想情况下，智能体将学会通过以下方式应对这些变化：提前派遣电动垂直起降飞行器前往目的地，以弥补风力影响造成的速度减慢；如果充电设备发生故障，则让电动垂直起降飞行器额外充电一步；机械故障排除后立即起飞，以节省时间。在此过程中，智能体仍应确保乘客安全性达到最高水平，因此发生碰撞将受到严厉惩罚。每个案例研究中，智能体将接受 50 个回合（ 72000 步）的测试，每个回合代表垂直起降机场运行的 24 小时，即 1 天。随后，将其与随机智能体和先到先服务（FCFS）智能体进行比较。随机智能体的设计相当简单，每一步都会随机采取一个动作，而不考虑状态空间。另一方面，先到先服务智能体将根据队列来决定哪架电动垂直起降飞行器应当起飞，哪架应当降落并充电。电动垂直起降飞行器在队列中等待时，会停在普通停机位；如果普通停机位已满，则会停在悬停点。先到先服务智能体还会为每架电动垂直起降飞行器充电，直到电量达到或超过 60%，之后便让其起飞并进入降落队列。

A. 学习曲线

GRL智能体经过了 30 万步（即 200 个回合）的训练，训练图见图7。从奖励和损失图中可以看出，该智能体能够泛化环境，因为奖励稳步增加，而损失稳步下降。在约第 60 个回合时，损失突然增加，这表明智能体未能根据接收到的状态空间信息做出恰当的近似；不过，损失最终回落并继续下降。此外，图7还展示了其他图表：每个无人机在每个回合中的平均电池电量；智能体每个回合完成的良好起飞和良好降落次数；单架无人机每个回合的平均延误（小时）；以及智能体每个回合经历的碰撞次数。这些信息是在训练期间记录的，有助于我们解读智能体的表现。我们可以看到，智能体起初的电池电量较低，但最终学会始终将电量保持在 30% 以上，因为低于这一水平会导致奖励惩罚。延误起初也很严重，但随着智能体逐渐熟悉起飞和降落，延误也开始减少。遗憾的是，智能体始终没能学会避免所有碰撞：它虽能避开一些碰撞，但随着时间推移，发生的碰撞反而更多。这是因为安全权重不够高。如果安全奖励权重更高，智能体就会被迫尽量减少所有碰撞以维持较高奖励，而不是试图通过完成更多良好的起飞和降落来抵消碰撞惩罚。

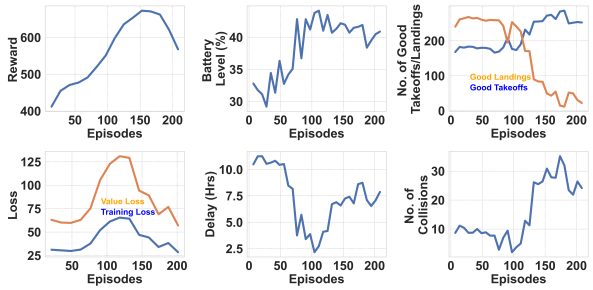

图7：GRL智能体的奖励和损失曲线，以及用于解释模型的环境指标

B. 案例研究

在案例研究中，我们选择了在16万步（即111个回合）时训练完成的理想GRL模型。训练到这一阶段时，碰撞次数最低，成功起飞和降落的次数较为均衡，延误接近局部最小值，平均电池电量则接近局部最大值，如图7所示。案例研究结果见图8。

案例1： GRL智能体在各项指标上的表现都很好，唯独碰撞次数除外。值得注意的是，GRL智能体在各项指标上的波动都更大，这可能是因为智能体在每个回合中都会适应不同情况。如果能有更多训练时间来微调其策略，这些波动就会更小、更可控。先到先服务智能体的表现排名第二，且各项指标的表现都更稳定。随机智能体的表现是三者中最差的，这符合预期。

案例2： GRL智能体和先到先服务智能体在某些指标上的表现大幅下降，在其他指标上的下降则较小。正如预期，GRL智能体会对环境中的噪声扰动作出反应：奖励下降，表明要在所有指标上实现性能最大化存在一定困难；电池电量下降了13%，这直接导致充电端口没有发生故障；延误也有所增加，同时良好起飞的次数减少，这是强风和电动垂直起降飞行器在起飞过程中发生机械故障所致；良好降落的次数也有所减少，因为电动垂直起降飞行器无法再准时降落；最后，碰撞次数略有减少，这与电动垂直起降飞行器起飞次数减少直接相关。先到先服务智能体也因相同原因表现下降，而随机智能体在除碰撞之外的所有指标上都远落后于另外两个基线。有趣的是，在存在噪声时，先到先服务智能体的总体延误更少，这可能是因为它在完成任务时（让电动垂直起降飞行器起飞，再让其在两个队列中降落）表现一致，从而不断为电动垂直起降飞行器安排新的时间表并重置其延误。如果电动垂直起降飞行器在队列中但不在队首，先到先服务智能体会跳过它，这样可以节省时间并减少延误。另一方面，GRL智能体会等待电动垂直起降飞行器进入其空域后才对其采取行动，这可能导致延误时间看似增加。

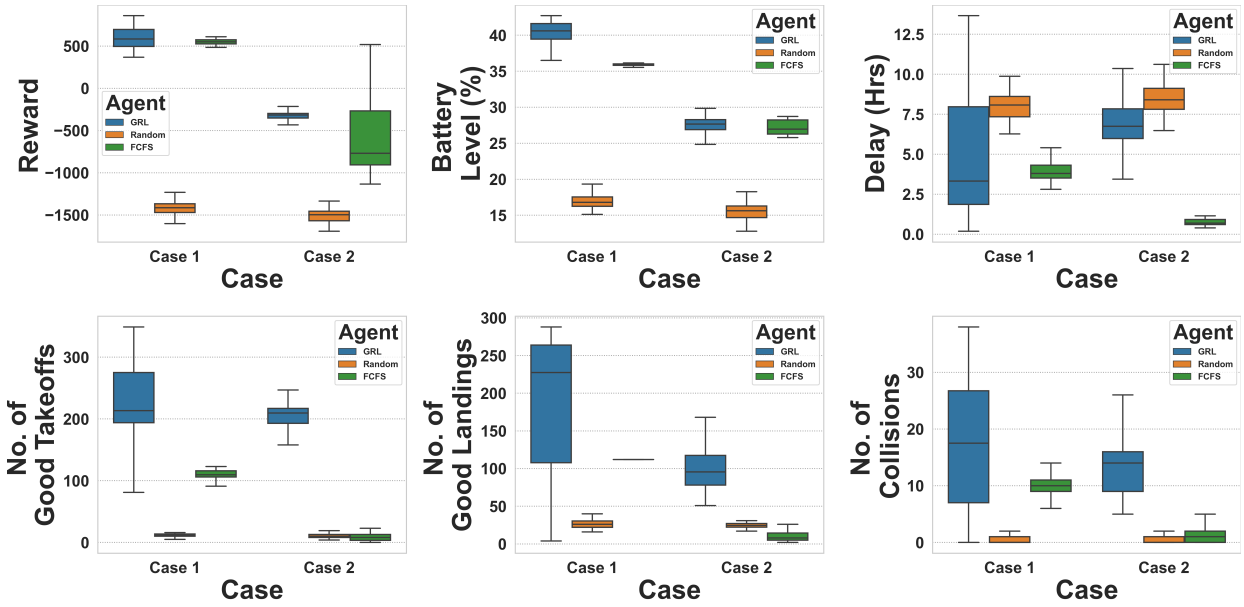

图8：案例研究结果。案例1中，三个智能体在无噪声环境下进行测试；案例2中，它们在有噪声的环境下进行测试。

这些结果表明，GRL智能体确实能够学习我们仿真环境的参数空间，并且能够以不同程度的成功适应仿真中添加的噪声。如果有更多时间进行训练和调整超参数，并采用更合适的奖励权重组合，我们相信无论环境中是否存在噪声，该智能体都能在每项指标上取得良好表现。

C. 动作分析

此外，我们记录了仿真环境中有噪声和无噪声两种情况下的 GRL智能体，各运行 5 个回合，并在图 9 中绘制了动作分布。该图展示了智能体在这 5 个回合中选择各个动作的次数，以及选择这些动作的时间占比。我们马上注意到，智能体并未使用分配给它的所有动作，这可能意味着它认为这些动作对于最大化奖励并非必要，也可能是因为它还没有足够的时间学会如何正确使用这些动作。智能体可能无需使用所有动作的另一个原因，是它所控制的电动垂直起降飞行器数量较少。值得关注的是，随着规模扩大，智能体将如何表现，以及它如何利用多个停机位和多个悬停点来维持当前的运行节奏。另一个重要的观察是带噪声时智能体的动作分布。虽然选择的动作相同，但动作频率发生了变化，尤其是避免碰撞和继续这两个动作。这表明智能体尝试实时调整，以适应噪声；它可能试图避免因恶劣风况而导致的碰撞，或应对类似情况。这也体现了我们所选模拟器的有效性。我们生成的风矢量足够逼真，给智能体带来了真正的挑战，也因此让它得以进一步泛化环境以及因加入噪声而扩大的参数空间。

VI. 结论

本文针对存在多种环境不确定性（阵风效应）和观测不确定性（电池接口故障以及电动垂直起降飞行器起飞延误）的城市空中交通垂直起降机场运营管理问题，提出了一种基于图的强化学习方法。我们方法的创新之处在于将图神经网络引入强化学习框架，用于对状态空间进行特征抽象。针对马尔可夫决策过程设计的奖励函数，是由若干项加权求和构成，用于量化良好的起飞、降落、电池状态、延误和安全性。首先，我们将城市空中交通垂直起降机场运营管理问题建模为马尔可夫决策过程，并采用策略梯度强化学习方法（近端策略优化）求解。基于图神经网络的策略网络由两个图卷积网络组成，分别用于编码垂直起降机场状态信息和电动垂直起降飞行器状态信息（两者均以图表示）。我们使用 AirSim 和 Unreal 引擎，将强化学习环境建模为基于 OpenAI Gym 的环境。我们在两组测试场景中测试了训练得到的策略（性能最佳者）。第一组测试场景与训练场景来自同一分布，第二组测试场景则存在不确定性。作为基准比较，我们采用了先到先服务策略和随机智能体。两组场景均包含 50 个未见过的回合。在这两种情况下，与其他方法相比，GRL智能体在总奖励方面均具有明显优势。进一步查看构成奖励的各项指标可以发现，GRL智能体仅在碰撞次数这一指标上的表现较差。

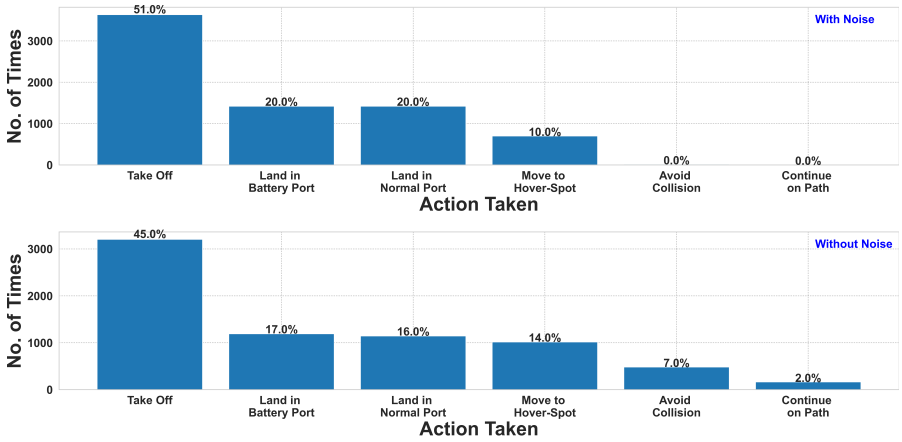

图 9：动作分布　　在有噪声和无噪声环境中，GRL 智能体在 5 个回合中累积的

训练时提高碰撞次数对应的权重并优先考虑碰撞次数，可以缓解这一问题。这两项案例研究展示了 GRL智能体在未见过的场景及存在不确定性的情况下进行泛化的能力。对所采取动作的进一步分析表明，不确定性如何改变决策，使智能体减少起飞、增加避碰，因此采取更为保守的策略。

参考文献

[1] “联合国称，到 2050 年，预计全球 68% 的人口将居住在城市区域——联合国经济和社会事务部”，2018 年 5 月。 [在线]。网址：https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html [2] R. Rothfeld、A. Straubinger、M. Fu、C. Al Haddad 和 C. Antoniou，“城市空中交通”，载于新兴交通系统需求。爱思唯尔，2020 年，第 267–284 页。[3] A. Bauranov 和 J. Rakas，“面向城市空中交通的空域设计：概念与方法综述”，《航空航天科学进展》，第 125 卷，文章编号 100726，2021 年。 [在线]。网址：https://www.sciencedirect.com/science/article/pii/S0376042121000312 [4] T. A. HORNE，“接下来：电动垂直起降飞行器：Uber 引领迈向喷气机时代般的未来”，《AOPA飞行员》，2019 年。[5] P. D. Vascik 和 R. J. Hansman，“通过洛杉矶探索性案例研究识别航空按需出行的约束”，载于第 *17* 届 *AIAA*航空技术、集成与运行会议，2017 年，第 3083 页。[6] M. Daskilewicz、B. German、M. Warren、L. A. Garrow、S.-S. Boddupalli 和 T. H. Douthat，“垂直起降机场选址进展及电动垂直起降飞行器日常通勤航程需求估算”，载于*2018* 年航空技术、集成与运行会议，2018 年，第 2884 页。[7] N. M. Guerreiro、G. E. Hagen、J. M. Maddalon 和 R. W. Butler，“采用先到先服务垂直起降机场调度算法的城市空中交通垂直起降机场容量与吞吐量”，载于*AIAA* 航空 *2020* 论坛，2020 年，第 2903 页。[8] L. Preis 和 M. Hornung，“垂直起降机场运行建模、基于智能体的仿真与参数值设定”，《电子学》，第 11 卷，第 7 期，文章编号 1071，2022 年。[9] A. Bacchini 和 E. Cestino，“电动垂直起降飞行器构型比较”，《航空航天》，第 6 卷，第 3 期，文章编号 26，2019 年。[10] Y. Li，“深度强化学习：综述”，*arXiv* 预印本 *arXiv:1701.07274*，2017 年。[11] K. Zhang、Z. Yang 和 T. Bas.ar，“多智能体强化学习：理论与算法精选综述”，2019 年。 [在线]。网址：https://arxiv.org/abs/1911 10635 。[12] W. Kool、H. Van Hoof 和 M. Welling，“注意力机制，学会解决路径规划问题！”，载于第七届国际学习表征会议（*ICLR 2019*），2019 年。

[13] T. D. Barrett、W. R. Clements、J. N. Foerster 和 A. I. Lvovsky，《使用强化学习进行探索式组合优化》，*arXiv* 预印本 *arXiv:1909.04063*，2019。[14] S. Paul 和 S. Chowdhury，《基于图的城市空中交通机队调度强化学习框架》。 [在线]。网址：https://arc.aiaa.org/doi/abs/10.2514/6.2022-3911[15] E. Khalil、H. Dai、Y. Zhang、B. Dilkina 和 L. Song，《在图上学习组合优化算法》，载于神经信息处理系统进展，2017 年，第 6348–6358 页。[16] Y. Kaempfer 和 L. Wolf，《使用置换不变池化网络学习多旅行商问题》，*ArXiv*，第 abs/1803.09621 卷，2018 年。[17] R. A. Jacob、S. Paul、W. Li、S. Chowdhury、Y. R. Gel 和 J. Zhang，《利用图上的强化学习重构不平衡配电网络》，载于*2022* 年 *IEEE* 德克萨斯电力与能源会议（*TPEC*），2022 年，第 1–6 页。[18] Z. Li、Q. Chen 和 V. Koltun，《利用图卷积网络和引导式树搜索进行组合优化》，载于神经信息处理系统进展，2018 年，第 539–548 页。[19] A. Nowak、S. Villar、A. S. Bandeira 和 J. Bruna，《关于使用图神经网络学习二次分配算法的简述》，*stat*，第 1050 卷，第 22 页，2017 年。[20] E. V. Tolstaya、J. Paulos、V. R. Kumar 和 A. Ribeiro，《使用空间图神经网络进行多机器人覆盖与探索》，*ArXiv*，第 abs/2011.01119 卷，2020 年。[21] Q. Sykora、M. Ren 和 R. Urtasun，《多智能体路由价值迭代网络》，载于第 *37* 届国际机器学习会议（*ICML 2020*），2020 年。[22]S. Paul 和 S. Chowdhury，《利用胶囊网络和注意力机制解决容量约束车辆路径问题的可扩展图学习方法》，载于国际设计工程技术会议与计算机和信息工程会议，第 3B 卷：第 48 届设计自动化会议（DAC），2022 年 8 月，v03BT03A045。 [在线]。网址：https://doi.org/10.1115/DETC2022-90123[23] H. Dai、E. B. Khalil、Y. Zhang、B. Dilkina 和 L. Song，《在图上学习组合优化算法》，载于神经信息处理系统进展，2017 年。[24] S. Paul、P. Ghassemi 和 S. Chowdhury，《利用胶囊注意力网络在图上为多机器人任务分配学习可扩展策略》，载于*2022* 年国际机器人与自动化会议（*ICRA*），2022 年，第 8815–8822 页。[25] P. Hernandez-Leal、B. Kartal 和 M. E. Taylor，《多智能体深度强化学习是答案还是问题？简要综述》，学习，第 21 卷，第 22 页，2018 年。[26] A. Behjat、H. Manjunatha、P. K. Kumar、A. Jani、L. Collins、P. Ghassemi、J. Distefano、D. Doermann、K. Dantu、E. Esfahani等，《在复杂对抗环境中学习机器人群体战术》，载于*2021* 年国际多机器人与多智能体系统研讨会（*MRS*）。IEEE，2021 年，第 83–91 页。

[27] X. Lyu、Y. Xiao、B. Daley 和 C. Amato，《多智能体强化学习中集中式与去中心化评论家的对比》，*arXiv* 预印本 *arXiv:2102.04402*，2021 年。[28] J. Schulman、F. Wolski、P. Dhariwal、A. Radford 和 O. Klimov，《近端策略优化算法》，*arXiv* 预印本 *arXiv:1707 06347*，2017 年。。[29] J. Schulman、S. Levine、P. Moritz、M. I. Jordan 和 P. Abbeel，《信赖域策略优化》，2015 年。 [在线]。网址：https: //arxiv.org/abs/ 1502 05477 。[30] G. Brockman、V. Cheung、L. Pettersson、J. Schneider、J. Schulman、J. Tang 和 W. Zaremba，《OpenAI Gym》，2016 年。[31] A. Hill、A. Raffin、M. Ernestus、A. Gleave、A. Kanervisto、R. Traore、P. Dhariwal、C. Hesse、O. Klimov、A. Nichol、M. Plappert、A. Radford、J. Schulman、S. Sidor 和 Y. Wu，《Stable Baselines》，https://github.com/hill-a/stable-baselines，2018 年。[32] A. Ilyas、L. Engstrom、S. Santurkar、D. Tsipras、F. Janoos、L. Rudolph 和 A. Madry，《深度策略梯度算法真的是策略梯度算法吗？》，*CoRR*，第 abs/1811.02553 卷，2018 年。 [在线]。网址：http://arxiv.org/abs/ 1811 02553 。W. Kool、H. van Hoof 和 M. Welling，《注意，学会解决路径规划问题！》，载于国际学习表征会议，2019 年。 [在线]。网址：https://openreview.net/ forum?id= ByxBFsRqYm [33] [34] S. Zhang、H. Tong、J. Xu 和 R. Maciejewski，《图卷积网络：全面综述》，*Computational Social Networks*，第 6 卷，第 1 期，第 1–23 页，2019 年。[35] S. Shah、D. Dey、C. Lovett 和 A. Kapoor，《AirSim：用于自动驾驶车辆的高保真视觉与物理仿真》，载于*Field and Service Robotics*。Springer，2018 年，第 621–635 页。[36] A. Sanders，《虚幻引擎 4 入门》。AK Peters/CRC Press，2016 年。[37] G. Brockman、V. Cheung、L. Pettersson、J. Schneider、J. Schulman、J. Tang 和 W. Zaremba，《OpenAI Gym》，*arXiv* 预印本 *arXiv:1606.01540*，2016 年。