

基于未知场景的无人机路径规划算法

常文文^{1,2} 邓 锴^{1,2} 普 阳^{1,2} 荆 鑫^{1,2}

CHANG Wenwen DENG Kai PU Yang JING Xin

摘 要

针对无人机在危险未知环境中路径规划效果欠佳的问题,文章提出了一种基于改进神经符号智能搜索算法的无人机路径规划方案。为解决无人机在大型危险环境中受限时效性的路径规划难题,通过组合神经符号系统,集成神经符号视觉感知、推理及基础模块来处理原始感官输入,实现了无人机在现实场景中可解释的路径规划;引入因果表征解缠学习方法,能够精准识别表征中的因果与非因果因素,有效提升了无人机在低可见度场景下的路径规划性能;将提取的因果因素传递至后续策略学习过程,消除非因果因素的干扰,进一步增强了模型的泛化能力。在 AirSim 仿真平台上的实验结果表明,与原有方法相比,改进算法的路径规划成功率提升了 7.8%。

关键词

无人机;神经符号系统;未知环境;路径规划

doi: 10.3969/j.issn.1672-9528.2025.08.013

0 引言

无人机导航技术近年来发展迅猛,已成为机器人学和人工智能领域的研究重点,引起了学术界和工业界的广泛关注。该技术为精准农业^[1-2]、搜索与救援^[3-4]、野生动物保护^[5]等多个应用场景提供了灵活、高效且经济的解决方案。随着无人机路径规划技术的不断发展,其在情报、监视与侦察(ISR)任务中的应用越来越广泛。无人机相比传统方法,具有显著的优势,尤其是在覆盖广阔区域和进入高风险区域方面,能够有效降低对人员生命安全的威胁^[6]。然而,无人机在复杂且不确定的环境中完全自主执行任务依然面临重大挑战。目前,由于许多 ISR 任务主要在不可预测且高度复杂的环境中进行,任务执行的不确定性增加了。此外,如何在任务描述简短且信息不完全的情况下基于机载传感器进行实时决策,并设计能够自主执行任务的无人机系统,依然是技术研究中的难点。

近年来,随着大型多模态模型(large multimodal models, LMM)^[7-9]技术的迅速发展,这些模型在多种机器人任务中已广泛应用。然而,LMM 主要基于多样化的大型数据集进行训练,而单一的端到端模型消耗的资源巨大。但是,此模型在失败时缺乏可解释性,并且可扩展性较差,尤其在对抗性环境下的性能表现更不佳。此外,LMM 缺乏准确的视觉

信息和对真实世界的状态表示。因此,不适合无约束环境中的搜索问题。

目前的搜索方法主要是基于(state-of-the-art, SoTA)视觉和多模态语言模型来感知环境执行飞行任务^[10-12]。将搜索问题先进行抽象化处理,然后和基于模型的规划器^[13-15]集成,从而使其有更好的推广性、稳健性和效率。然而,这种方法缺乏明确的视觉推理方法和持久的环境模型,限制了对环境的符号化表示和做出明智决策的能力。

为解决这些限制,本文首先引入了神经符号智能搜索算法,基于分层模型的规划组件,其利用高级规划器确定整体搜索策略,利用中级规划导航到分配区域,使用低级规划在分配区域内有效搜索并避开障碍物;然后从因果关系的角度解决基于深度强化学习(deep reinforcement learning, DRL)的无人机防撞系统中的泛化问题,并通过因果解缠框架学习鲁棒和因果视觉表示;设计了背景干预模块有效地干扰了背景,有助于分离因果因素与非因果因素。

1 问题描述与模型建立

本文采用集中训练和分布式执行的方式训练策略。无人机仅能通过前置摄像头进行部分环境感知来避免碰撞。因此,本文设计了部分分散的可观测马尔可夫决策过程,以此增强对环境的感知能力。

1.1 观测空间

无人机的观测包括 3 个部分:对深度图像的表示 o^d 、目标在机身坐标系内的相对位置 o^r ,以及无人机当前时刻的速度 o^v 。

1. 郑州西亚斯学院计算机与软件工程学院 河南郑州 451100

2. 郑州西亚斯学院计算机工程实践中心 河南郑州 451100

[基金项目] 郑州西亚斯学院 2024 年度科研资助项目 (2024XKE116)

1.2 动作空间

本文利用连续的动作空间增强无人机飞行的多样性和可控性。可以从策略网络 $\pi(s)$ 中计算得出动作函数：

$$a = [v_x^{\text{cmd}}, v_z^{\text{cmd}}, v_\omega^{\text{cmd}}] \quad (1)$$

式中： v_x^{cmd} 为无人机的前进速度，表示其沿无人机机身坐标系前轴的速度； v_z^{cmd} 为爬升速度，表示无人机的垂直运动； v_ω^{cmd} 为转向速度，表示无人机绕其垂直轴的旋转运动。

1.3 奖励函数

每个代理选择合适的奖励函数为其提供更精确的反馈，奖励函数由两部分组成，一部分 r_g 引导无人机朝目标飞行；另一部分 r_c 引导无人机避开障碍物。具体奖励函数为：

$$r = r_g + r_c \quad (2)$$

r_g 具体表达式为：

$$r_g = \begin{cases} r_{\text{arrived}} & \text{if } d_t < 0.5 \\ \alpha_{\text{goal}} \cdot (d_t - d_{t-1}) & \text{其他} \end{cases} \quad (3)$$

式中： d_t 为 t 时刻无人机与目标点之间的距离。

$$r_c = \begin{cases} r_{\text{collision}} & \text{if crash} \\ \alpha_{\text{avoid}} \cdot \max(d_{\text{safe}} - d_{\text{min}}, 0) & \text{otherwise} \end{cases} \quad (4)$$

设参数的值分别为 $r_{\text{arrival}} = 50$, $r_{\text{collision}} = -10$, $\alpha_{\text{goal}} = 3$, $\alpha_{\text{avoid}} = -0.05$ 与 $d_{\text{safe}} = 5$ 。

2 系统架构及网络结构

2.1 网络架构

本文设计了一个基于变分自动编码器的框架进行视觉提取和策略学习。如图 1 所示，本框架遵循 (soft actor-critic, SAC) 范式，输入为深度图像、当前速度和相对目标位置，输出为飞行控制动作。本文提出了一种因果表示解耦方法来优化视觉表示提取，并将因果传递给后续的策略学习。

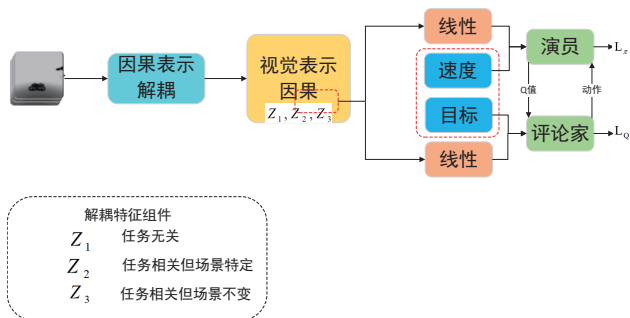


图 1 无人机防撞架构

本网络优化表示提取过程主要通过因果表示解耦来优化表示提取过程的，此过程能发现隐藏的因果因素并过滤掉非因果因素，从而提高深度强化学习模型的泛化能力。本文利用特征解耦中的典型做法，将提取的表示分为 3 个典型成分，分别为：与任务无关的成分 $Z_1 \in \mathbf{R}^{n_1}$ ；与任务相关但特定于场景的成分 $Z_2 \in \mathbf{R}^{n_2}$ ；与任务相关且与场景不变的成分

$Z_3 \in \mathbf{R}^{n_3}$ 。障碍物分布 Z_2 和障碍物距离 Z_3 可以为防撞任务提供关键信息的因果因素。与此不同的是， Z_1 为背景模式，与任务无关，而且由于数据驱动的训练而可能构建虚假相关性，导致了无人机系统在看不见的场景中部署时泛化能力降低。本文提出的因果表示解耦法可以对以上 3 种成分进行识别，并将 Z_2 和 Z_3 作为后续的策略学习，从而明确消除 Z_1 的影响。因果表示解耦的关键是引导表示组件按预期学习不同的语义概念。因果表示解耦的具体过程如图 2 所示，本文使用变分自动编码器进行表征提取，然后设定特定的损失函数引导预定义组件学习不同的语义概念。

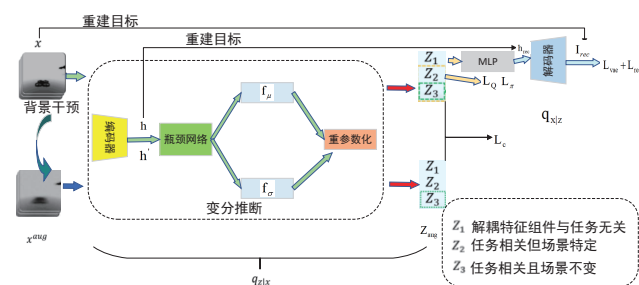


图 2 因果表征解耦的图示

2.2 神经符号系统

神经符号系统主要由 3 部分组成：神经符号视觉感知、推理和基础组件 (GRiD)、符号环境模型和符号分层规划组件 (SNaC)，如图 3 所示，无人机在模拟环境 Airsim 中运行，并配备了 RGB 摄像头、深度摄像头和 GPS 传感器。系统包括三维感知、接地、推理 (GRiD) 组件利用代码生成器和视觉基础模型处理传感器数据。预测的生成主要包括神经分割、对象检测、属性分类以及符号二维跟踪器和三维投影仪。将预测发送到能够维护信息图并生成目标报告的环境模型。分层计划包括选择、导航和覆盖 (SNaC)，其中兴趣区域 (area of interest, AOI) 选择、AOI 导航和区域覆盖模块分别生成高级、中级和低级规划。

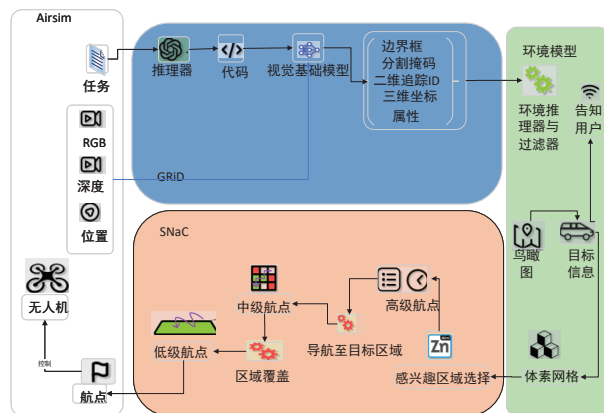


图 3 神经符号系统框架图

兴趣边缘 (edge of interest, EOI) 由描述和位于不同 AOI 内概率的组合来指定。通过禁区 (keep-out zone, KOZ) 增加

进一步的限制，禁区为无人机不得进入的区域。将基准测试的时间限制为 5 min，环境中具有相关类型和颜色的汽车。EOI 选择制定在规定的不会产生歧义的唯一描述的 AOI 内。无人机优先关注最有希望的 AOI，并合理地分配时间。

3 实验与结果分析

3.1 实验与参数设置

本文利用虚幻引擎 (unreal engine, UE) 和 Airsim 模拟器开发的几个虚拟环境运行实验。所有模拟实验均运行在 Ubuntu 20.04 的系统上，该系统配备 Intel i7 处理器和 NVIDIA GeForce RTX 3060 GPU。在每一轮模拟中，无人机被分配了三维空间内随机选择的起始位置和目标点。此设计能使无人机能彻底探索高维观测空间，增强了学习策略的鲁棒性，表 1 为本实验的超参数。

表 1 策略训练的超参数

参数名称	值
重放缓冲区 β 容量	10 000
批处理大小	128
最大回合数	150
更新次数	200
折扣因子 γ	0.5
优化器	Adam
编码器学习率	10^{-4}
评论家学习率	10^{-4}
评论家目标跟踪率	1
评论家 Q 函数软更新率 τ_Q	0.005
评论家编码器软更新率 τ_{enc}	0.025
演员学习率	10^{-2}
演员更新频率	1
演员对数标准差界限	$[-5, 1]$

3.2 性能指标和实验场景

本文为了评估不同的性能，采用了以下评估指标：

成功率：在规定时间内到达既定目标点且未发生任何碰撞的代理百分比。

按路径长度加权的成功率 (success weighted by path length, SPL)：测量无人机能否通过最短路径成功到达目的地。

$$SPL = \frac{\sigma \cdot l}{\max(l, p)} \quad (5)$$

式中： l 为无人机从出发点 to 目标点的最短路径距离； p 为无人机实际行驶的路径长度； σ 为成功或失败的二进制指标。

额外距离：无人机行程与出发点和目标点之间最短距离的平均额外距离。

平均速度：测试过程中无人机的平均速度。



图 4 模型训练与测试的模拟场景

图 4 为模型训练与测试的模拟场景，其场景分别为方块、城市、乡村场景，本文利用不同的场景来评估深度强化学习模型的泛化能力。其中一个用于训练，其他用于测试。

3.3 性能评估

本文在两种初始化设置的 3 个看不见的测试场景中测试本文和之前的 SOTA 方法，即 SAC+RAE，本文将无人机的起始点和目标点分别进行两种不同的初始化模式，一种为随机模式，即无人机的起点和目标位置均在一个长宽 16 m、高 4 m 的立方体区域内随机生成的；另一种为将无人机定位在一个特定高度，且半径为 12 m 的圆上，如表 2 和表 3 所示。

表 2 ABLE II 性能与 SOTA 方法对比

场景	方法	成功率 /%	SPL /%	额外距离 /m	平均速度 / (m·s ⁻¹)
方块	SAC+RAE	85.4	73.8	1.511/1.282	0.895/0.132
	本文方法	88.2	81.9	0.721/0.947	0.987/0.101
城市	SAC+RAE	80.2	66.3	2.087/1.681	1.811/0.174
	本文方法	88.0	67.5	3.062/2.653	0.958/0.072
乡村	SAC+RAE	72.7	59.9	2.121/1.711	0.802/0.170
	本文方法	88.5	71.8	2.56/2.74	0.98/0.117

表 3 在圆圈初始化下与 SOTA 方法的性能 (平均值 / 标准差) 比较

场景	方法	成功率 /%	SPL /%	额外距离 /m	平均速度 / (m·s ⁻¹)
方块	SAC+RAE	53.4	49.7	1.765/0.507	0.800/0.045
	本文方法	98.3	92.3	1.545/0.331	1.118/0.054
城市	SAC+RAE	72.0	67.4	1.651/0.878	0.955/0.137
	本文方法	99.1	91.7	1.688/0.883	1.042/0.061
乡村	SAC+RAE	40.5	31.6	6.673/1.292	0.966/0.137
	本文方法	99.0	70.6	9.951/3.823	1.005/0.050

从结果可以看出，本文的算法可以在更高的平均速度下显著提高导航成功率和 SPL，说明因果表示解缠能有效提高深度强化学习模型在看不见的场景中的泛化能力。其次，本算法规划的飞行路径略长于 SAC+RAE，因为本文的路径

规划包含更多的防撞动作，而 SAC+RAE 规划的路径由于看不见背景干扰而导致更多碰撞。因此，本文方法能在更高的平均速度下显著提高导航成功率和 SPL，并且能有效地提高 DRL 模型在看不见的场景中的泛化能力。其次，本方法规划的路径略长于 SAC+RAE，包含更多的防撞动作，而 SAC+RAE 规划的路径由于看不见的背景干扰而导致更多的碰撞。

3.4 消融实验

本文将任务拆分为 3 个部分，即 Z_1 、 Z_2 和 Z_3 ，并且仅使用 Z_2 和 Z_3 作为任务相关组件进行后续策略学习，消除了与任务无关的组件 Z_1 的影响。通过对不同组件进行消融实验，结果如表 4 所示，由结果可得， Z_2 和 Z_3 对于结果学习都很重要，两者结合的性能最佳，其次，使用 Z_1 导致泛化能力降低，因为 Z_1 包含与任务无关的信息，此信息容易构建虚假相关性并导致错误的动作预测。

表 4 消融实验

Z_1	Z_2	Z_3	成功率 /%
	√	√	88.5
	√		64.3
		√	60.3
√	√	√	71.0

4 结论

为解决基于 DRL 的无人机路径规划系统在未知场景中的泛化能力较差的问题，本文设计了一种新的表示学习方法，首先将神经符号感知与概率世界模型进行整合，能实现无人机高效的路径规划，然后，将视觉表示划分为具有特定语义概念的几个组件，最后，仅传递因果组件进行后续的策略学习，能够消除与组件无关组件的影响，提高了模型的泛化能力。

参考文献:

[1] 吕茂隆, 丁晨博, 韩浩然, 等. 基于深度强化学习的无人机自主感知-规划-控制策略[J]. 自动化学报, 2025, 51(6): 1305-1319.

[2] 候磊, 贾贝熙, 杜子亮, 等. 深度强化学习在无人机智能控制中的应用研究[J]. 战术导弹技术, 2024(6):107-117.

[3] 唐嘉宁, 李成阳, 周思达, 等. 基于深度强化学习的无人机自主探索方法[J]. 计算机科学, 2024, 51(S2):154-159.

[4] 梁秀满, 刘子良, 刘振东. 基于深度强化学习的改进 RRT 算法路径规划[J/OL]. 系统仿真学报, 1-12[2025-05-05]. <https://doi.org/10.16182/j.issn1004731x.joss.24-0494>.

[5] 王娜, 马利民, 姜云春, 等. 基于多 Agent 深度强化学习的无人机协作规划方法[J]. 计算机应用与软件, 2024, 41(9):83-89.

[6] BO L, ZHANG T Z, ZHANG H X, et al. 3D UAV path planning in unknown environment: a transfer reinforcement

learning method based on low-rank adaption[J]. Advanced engineering informatics, 2024, 62:102920.

[7] GUO J R, ZHOU G Z, HUANG H L, et al. Advancements in UAV path planning: a deep reinforcement learning approach with soft actor-critic for enhanced navigation[J]. Unmanned systems, 2024, 13(5):1065-1084.

[8] SUN G Y, XU Q, ZHANG G Y, et al. An intelligent UAV path-planning method based on the theory of the three-dimensional subdivision of earth space[J]. ISPRS international journal of geo-information, 2023, 12(10):397.

[9] LÜ H, CHEN Y D, LI S B, et al. Improve exploration in deep reinforcement learning for UAV path planning using state and action entropy[J]. Measurement science and technology, 2024, 35(5): 56206.

[10] SABZEKAR S, SAMADZAD M, MEHDITABRIZI A, et al. A deep reinforcement learning approach for UAV path planning incorporating vehicle dynamics with acceleration control[J]. Unmanned systems, 2023, 12(3):477-498.

[11] ZHOU Z Y, LIU M X, GUO J S, et al. End-to-end deep learning method with disparity correction for stereo matching[J]. Arabian journal for science and engineering, 2023, 49: 3331-3345.

[12] YANG Y, ZHONG C, RUAN Z S, et al. UAV path planning based on deep reinforcement learning[C]//8th International Conference on Computing, Control and Industrial Engineering (CCIE2024). Berlin: Springer, 2024: 46-53.

[13] LEE M H, MOON J. Deep reinforcement learning-based model-free path planning and collision avoidance for UAVs: a soft actor-critic with hindsight experience replay approach[J]. ICT express, 2023, 9(3): 403-408.

[14] LI B H, WU Y J. Path planning for UAV ground target tracking via deep reinforcement learning[J]. IEEE access, 2020, 8: 29064-29074.

[15] TANG J, LIANG Y G, LI K B. Dynamic scene path planning of UAVs based on deep reinforcement learning[J]. Drones, 2024, 8(2): 60.

【作者简介】

常文文(1999—), 女, 河南濮阳人, 硕士研究生, 助教, 研究方向: 计算机视觉、人工智能。

邓锴(1997—), 男, 河南南阳人, 硕士研究生, 助教, 研究方向: 算力网络、未来网络。

普阳(1990—), 男, 河南周口人, 硕士研究生, 助教, 研究方向: 计算机视觉、人工智能。

荆鑫(2003—), 男, 河南三门峡人, 本科, 研究方向: 计算机视觉、人工智能。

(收稿日期: 2025-04-10 修回日期: 2025-08-05)