无人机集群协同主动搜索的强化学习策略研究

肖子健[1]，夏晨钧[1]，徐杨罡[1]，任纪媛[1]，陈鑫磊[1, 2, 3]
（1. 清华大学深圳国际研究生院，广东 深圳 518000；2. 鹏城实验室，广东 深圳 518000；
3. RISC-V 国际开源实验室，广东 深圳 518000）

摘　要：在多变和复杂的灾害环境中，迅速定位幸存者是一项至关重要的任务，无人机（UAV, unmanned aerial vehicle）群的主动搜索能力在这一过程中发挥着关键作用。然而，无人机的传感器性能与其飞行高度紧密相关，覆盖范围和探测精度难以平衡。为了实现高效的搜索，无人机集群需要在高空飞行以覆盖更广的区域，同时在低空飞行以提高探测的准确性。此时，策略的制定对于无人机集群的协调和决策至关重要。为了应对这些挑战，提出了协同高度自适应强化学习（CARL, collaborative altitude-adaptive reinforcement learning）方法，该方法融合了可变高度传感器模型、基于信心的评估机制以及基于近端策略优化（PPO, proximal policy optimization）的高度自适应规划器。通过 CARL 方法，无人机能够根据实时情况动态地调整感知策略，并做出更加明智的决策。此外，引入了一种创新的奖励塑造策略，从而在广阔环境中最大化搜索效率。通过在多种条件下的模拟测试，CARL 方法在提高完全搜索率方面表现出色，相较于基线方法提升了 12%，充分证明了其在提升无人机集群在主动搜索任务中的有效性。

关键词：强化学习；贝叶斯学习；协同无人机集群；主动搜索框架
中图分类号：TP393
文献标志码：A
doi：10.11959/j.issn.2096−3750.2024.00413

Collaborative altitude-adaptive reinforcement learning for active search with unmanned aerial vehicle swarms

XIAO Zijian[1], Chen-Chun Hsia[1], XU Yanggang[1], REN Jiyuan[1], CHEN Xinlei[1, 2, 3]
1. Shenzhen International Graduate School, Tsinghua University, Shenzhen 518000, China
2. Pengcheng Laboratory, Shenzhen 518000, China
3. RISC-V International Open Source Laboratory, Shenzhen 518000, China

Abstract: Active search with unmanned aerial vehicle (UAV) swarms in cluttered and unpredictable environments poses a critical challenge in search and rescue missions, where the rapid localizations of survivors are of paramount importance, as the majority of urban disaster victims are surface casualties. However, the altitude-dependent sensor performance of UAV introduces a crucial trade-off between coverage and accuracy, significantly influencing the coordination and decision-making of UAV swarms. The optimal strategy has to strike a balance between exploring larger areas at higher altitudes and exploiting regions of high target probability at lower altitudes. To address these challenges, collaborative altitude-adaptive reinforcement learning (CARL) was proposed which incorporated an altitude-aware sensor model, a confidence-informed assessment module, and an altitude-adaptive planner based on proximal policy optimization (PPO) algorithms. CARL enabled UAV to dynamically adjust their sensing location and made informed decisions. Furthermore,

收稿日期：2024−09−08；**修回日期：**2024−09−20
通信作者：陈鑫磊，chen.xinlei@sz.tsinghua.edu.cn
基金项目：国家重点研发计划项目（No. 2022YFB3300703）；国家自然科学基金项目（No. 62371269）；深圳市稳定支持项目（No. WDZC20220811103500001）；清华大学深圳国际研究生院交叉科研创新基金项目（No. JC20220011）
Foundation Items: The National Key Research and Development Program of China (No. 2022YFC3300703), The National Natural Science Foundation of China (No. 62371269), Shenzhen 2022 Stabilization Support Program (No. WDZC20220811103500001), Tsinghua Shenzhen International Graduate School Cross Disciplinary Research and Innovation Fund Research Plan (No. JC20220011)

a tailored reward shaping strategy was introduced, which maximized search efficiency in extensive environments. Comprehensive simulations under diverse conditions demonstrate that CARL surpasses baseline methods, achieves a 12% improvement in full recovery rate, and showcase its potential for enhancing the effectiveness of UAV swarms in active search missions.

Key words: reinforcement learning, Bayesian learning, collaborative UAV swarms, active search framework

0　引言

主动搜索在各种任务中至关重要，例如搜救行动、精准农业、野生动物监测和边境安全，主要任务是在不同的环境中定位未知数量的目标。快速识别和定位目标的能力至关重要，尤其是在及时干预可以挽救生命的灾难场景中。研究表明，地震受害者的存活率在灾难发生72 h后急剧下降[1]，这凸显了及时干预的重要性。此外，Meera等[2]表明大约80%的城市灾害幸存者是地表受害者，因此，有效的地表主动搜索策略对于及时发现和营救幸存者至关重要。

无人机（UAV, unmanned aerial vehicle）因其敏捷性、空中优势视角和快速区域覆盖能力而特别适合主动搜索任务[3-5]。此外，无人机具有快速部署和适应各种地形的能力，这些特性显著提高了搜索效率。

然而，在大规模、三维环境中使用无人机集群进行主动搜索时[6-10]，飞行高度对目标检测模型性能的影响通常会阻碍集群的高效调度。例如，一项关于海拔高度对YOLO（you only look once）[11]目标检测器精度影响的研究表明，在10 m的高度，无人机集群检测性能最佳，但是视场（FOV, field of view）有限，而在较高的高度，由于噪声的增加和图像分辨率的降低，无人机集群检测精度显著下降[2]。这种检测精度与高度相关的特性对协调无人机集群实现高效的主动搜索任务提出了挑战。为了在有限的时间窗口内准确定位多个地面受害者，必须考虑高度变化引起的检测精度和视场之间的权衡。

关于主动搜索[12-28]的研究涉及各个方面，例如目标定位[14]、敏捷传感[13, 17]、集群协作[20, 24, 27-28]、路径规划[21-23]和噪声建模[12]。然而，海拔高度对三维空间检测性能的影响经常被忽视。同样，地面机器人的深度感知噪声模型并不关注在三维环境中运行的无人机所遇到的具体挑战。Igoe等[25]研究了高度对检测性能的影响，基于此，本文提出了一种算法，将这些考虑因素整合到无人机的决策过程中，利用强化学习来增强复杂三维环境中的搜索策略。

本文解决了在未知的大规模环境中无人机集群主动搜索的问题。本研究主要面临两个挑战：1) 环境的未知性和每架无人机观测的局部性带来的决策挑战，缺乏全面的信息很容易导致局部最优的搜索策略。2) 检测精度和FOV之间的权衡挑战，无人机必须在低检测精度与大FOV和高检测精度与小FOV之间进行权衡。

为了应对这些挑战，我们提出了协同高度自适应（CARL, collaborative altitude-adaptive reinforcement learning）方法，该方法是由两个关键模块组成的协作式高度自适应强化学习方法。CARL方法的第一个模块是置信度评估，该模块通过群体协调来解决未知环境和部分观察的挑战。具体来说，我们使用贝叶斯推理更新来自每个无人机和时间步长的时空传感数据，同时有效地减轻了传感器噪声对目标概率估计的影响。第二个模块是一个高度自适应规划器，该模块通过调整无人机的传感位置来动态平衡FOV和检测精度。该规划器使用基于近端策略优化（PPO, proximal policy optimization）的强化学习算法，并由基于熵的奖励塑造策略指导，该模块利用卷积神经网络-长短期记忆（CNN-LSTM, convolutional neural network and long short term memory）编码器从高维搜索空间中提取特征，提高了搜索效率。本文在不同环境条件下进行了模拟实验（最多涉及50个目标和10架无人机），实验结果表明，CARL方法的完全搜索率较基线算法提高了12%。

1　问题定义

本文的重点在于解决利用n架无人机主动搜索m个感兴趣目标的问题。这些无人机配备了YOLO[11]传感器，具有执行搜索任务的人体检测能力。每架无人机都拥有空中视角，可在大尺度环境中实现广泛地地面感知和覆盖。我们假设无人机可以通过

GPS准确确定自己的位置，每架无人机都通过不断移动和感知周围环境来检测和定位这些目标。多个无人机在未知环境中进行主动目标搜索情况如图1所示，其中描述的区域被解释为需要主动搜索的区域，底部以“X”标记的人像表示目标的位置。无人机下方的棱锥形阴影所覆盖的区域表示每架无人机的机载感知范围，绿色区域表示不在任何无人机观测范围内的未知区域。因此，每个无人机都必须根据给定时刻的当前状态确定下一步操作，以优化整个场景的整体主动搜索操作。

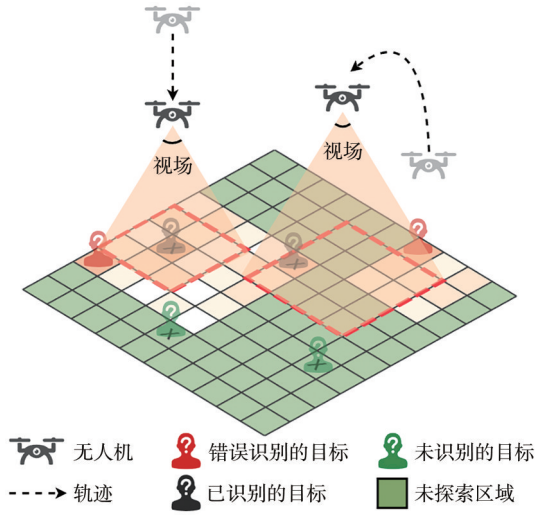


图1　多个无人机在未知环境中进行主动目标搜索情况

然而，无人机的机载传感器通常存在识别错误。在本文中，无人机利用视觉传感器捕获其视野内的物体，并采用YOLO物体检测算法来分析捕获的图像。该算法依靠置信度分数来识别对象，置信度分数随目标与传感器之间的距离而变化。距离传感器较远的物体通常不太可能被正确识别，这给无人机在观察过程中检测目标带来了不确定性。同时，视觉传感器的视野也随着无人机的高度而变化。

1.1　非定高传感器建模

影响无人机路径规划的关键因素是与机载传感器和识别算法相关的不确定性。当考虑传感器的噪声和识别算法产生的误差时，Meera等[2]分析了无人机在不同高度捕获的图像以及识别算法的性能，发现算法置信度的均值μ和方差σ^2与无人机的高度具有显著关系，在较高的高度，目标检测中出现假阴性或假阳性的可能性会增加。因此，本文使用了正态分布，并将目标识别算法的置信度得分建模为

$$\phi_p = \mathrm{clip}(\varphi_p, 0, 1), \varphi_p \sim \mathrm{N}(\mu(\eta_p, h), \sigma^2(h)) \quad (1)$$

其中，clip函数通过从正态分布中裁剪采样值来获得置信度分数，因为置信度分数在区间[0,1]之外是未定义的，η_p表示目标是否存在于位置p。无人机在过低的高度飞行会增加与障碍物相撞的风险，而飞得太高可能会导致严重的识别错误。因此，无人机飞行高度h被限制在一定的范围内。$\mu(\eta_p, h)$表示目标存在或不存在于位置p及高度h处的置信度得分均值。方差$\sigma^2(h)$包含了由于大气失真、运动模糊和目标图像分辨率降低而增加的噪声和不确定性[2]。

Meera等[2]探讨了传感器性能随高度变化的信息，本文整合了该研究的实验数据，并利用反比例函数$M(h)$拟合无目标条件下$\mu(\eta_p, h)$随高度h的变化，置信度分数的均值随传感器的高度变化如图2所示。$M(h)$的表达式为

$$M(h) = \frac{8.08}{h + 5.48} + 0.405 \quad (2)$$

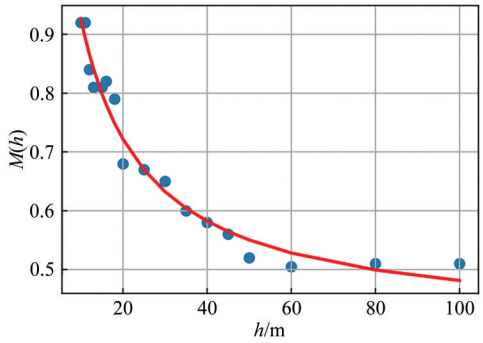


图2　置信度分数的均值随传感器的高度变化

置信度均值如式(3)所示，在这种建模方式下，目标位置的置信度均值随海拔高度的增加而减少，而非目标位置的置信度均值则随海拔高度的增加而增加。

$$\mu(\eta_p, h) = \begin{cases} M(h), \eta_p = 1 \\ 1 - M(h), \eta_p = 0 \end{cases} \quad (3)$$

1.2　问题建模

为了简化问题，本文进行了几个假设。首先，假设无人机仅在到达目标位置时执行目标感知，这是因为视觉传感器在捕捉移动物体方面通常表现不佳；假设无人机之间的通信始终可用，使它们能够共享信息并共同更新置信度图；另外，实际环境下传感器的检测性能通常还会受到无人机朝向、环境的气象因素等多方面的影响，为了聚焦问题，暂时只考虑无人机高度变化造成的影响。

主动搜索任务的一个目标是在给定当前环境状

态和每个时间步的置信度图$\boldsymbol{B}$的情况下，在下一个时间步为无人机提供操作，另一个目标是缩短总搜索时间。因此，目标函数定义为在有限时间内规划无人机集群的行动，以最大化置信度图的正信息增益。本文利用互信息量化无人机动作为环境中的目标置信度提供信息增益，置信度表示为I。对于环境中的特定位置p，当目标确实存在于该位置时，置信度分数接近1。相反，当没有目标时，置信度分数则较低。置信度计算方式为

$$I\left(\varphi_p;\tau\right)=\begin{cases}w_p\ln\dfrac{P(\varphi_p|\tau)}{P(\varphi_p)},\eta_p=1\\ w_e\ln\dfrac{1-P(\varphi_p|\tau)}{1-P(\varphi_p)},\eta_p=0\end{cases}\tag{4}$$

其中，$P(\varphi_p)$表示在时间t的位置p存在目标的概率，而$P(\varphi_p|\tau)$表示无人机集群采用动作轨迹$\tau=\{a_1,a_2,\cdots\}$并感知环境后，目标在位置p的概率。w_p和w_e表示两种类型的信息增益权重，由于在主动搜索任务中准确定位目标位置具有更大的意义，因此通常将权重设置为$w_p>w_e$。主动搜索任务的目标函数可以定义为

$$\tau^*=\arg\max_{\tau}\frac{\sum_{p,t}I(\varphi_p;\tau)}{T}$$
$$\text{s.t.}\quad h\in\left[h_{\text{Low}},h_{\text{High}}\right]\tag{5}$$

其中，T表示总搜索时间。

2　算法结构

CARL方法结构如图3所示，演示了CARL方法协助无人机在单个时间步长内做出决策，使无人机集群更好地完成主动搜索任务，以下将详细说明此过程的关键点。

2.1　置信度评估

置信度矩阵表示为$\boldsymbol{B}\in(0,1)^{L_1\times L_2}$（$L_1$，$L_2$表示区域网格数量），在初始时间步，置信度图中的每个网格单元（表示为$\boldsymbol{B}_0$）都被分配了一个基本概率值（0.5），表示关于目标存在的初始不确定性。理想情况下，完成无人机的主动搜索后，具有目标的网格的置信度分数应为1，而在没有目标的网格上，置信度分数应为0。置信度的更新策略参考了经典的贝叶斯推断原理

$$P(A|D)=\frac{P(D|A)\cdot P(A)}{P(D)}\tag{6}$$

其中，$P(A)$表示目标先验概率，代表未观测前对目标位置的判断，$P(D)$表示观测先验概率，代表观测发生的概率，$P(A|D)$为后验概率，代表观测对于目标位置判断的修正。据此，在无人机对特定区域进行重复搜索的场景下，本文将置信度更新规

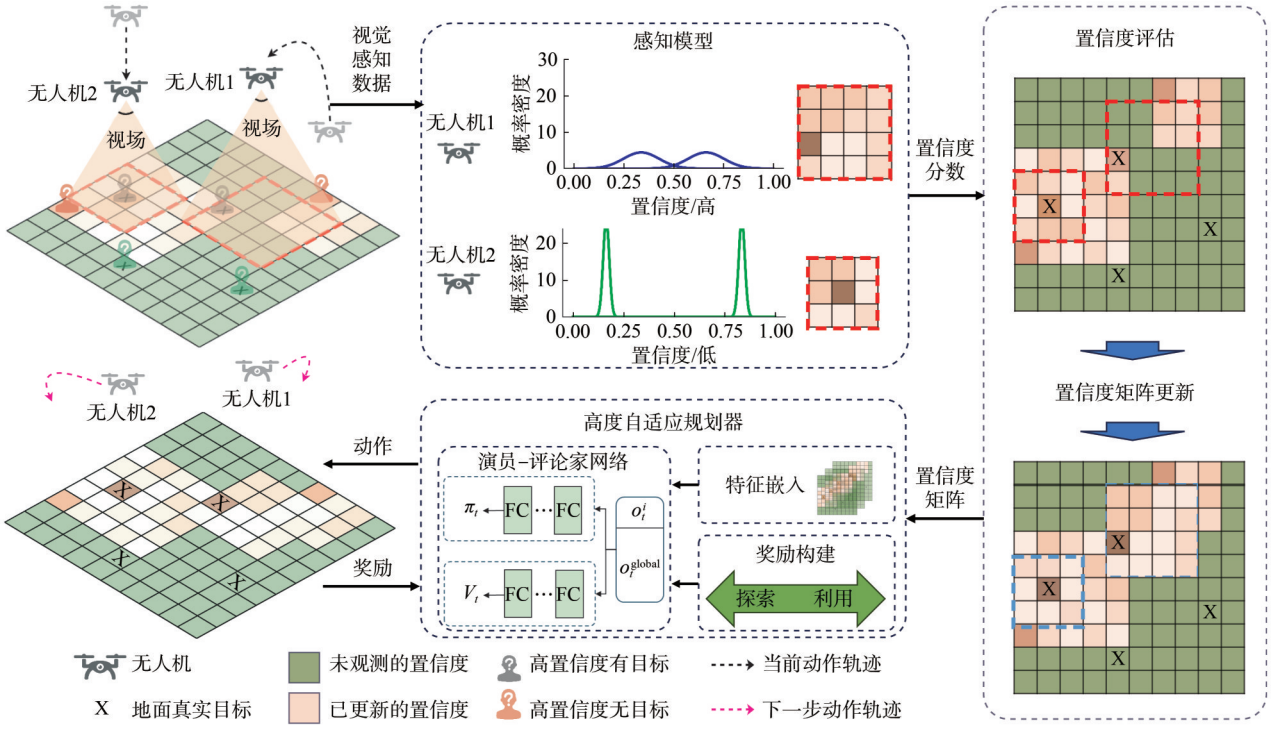


图3　CARL方法结构

则定义为

$$\boldsymbol{B}_t^{i+1}=\frac{\boldsymbol{\beta'}_t^{\,i}\circ\boldsymbol{B}_t^i}{(1-\boldsymbol{\beta'}_t^{\,i})\circ(1-\boldsymbol{B}_t^i)+\boldsymbol{\beta'}_t^{\,i}\circ\boldsymbol{B}_t^i},i\in(1,\cdots,n-1) \tag{7}$$

其中，“∘”表示元素积，$\boldsymbol{\beta'}_t^{\,i}\in(0,1)^{L_1\times L_2}$表示第$i$个无人机观测后的目标置信度分数矩阵，分数由第$i$个无人机在$t$时刻观测后增广得到。例如，在时间$t$，位于位置$p=(x,y)$的无人机$i$进行观测，得到$\boldsymbol{\beta}_t^i\in(0,1)^{L_{\text{fov}}\times L_{\text{fov}}}$（$L_{\text{fov}}$表示随无人机高度变化的视场的直径），然后将其通过边缘填充的方式扩充至$L_1\times L_2$（使用0.5填充，代表未观测）。

算法通过贝叶斯更新原理按顺序更新来自所有n个无人机的观测结果。在t+1时刻，置信度矩阵为$\boldsymbol{B}_{t+1}=\boldsymbol{B}_t^n$。该方法允许整合多个无人机观测，有效地提高了大规模搜索环境中置信度的准确性。

2.2　高度自适应规划器

1) 观测嵌入：$\boldsymbol{A}_t$表示UAV高度图，为一个形状与置信度相同的矩阵。如果网格未被任何UAV占用，则其值设置为$\boldsymbol{A}(x,y)=0$。对于占用的网格，该值对应于UAV的高度（$\boldsymbol{A}(x,y)=h_t^i$）。整体状态编码表示为这些部分的组合

$$\boldsymbol{S}_t=\{\boldsymbol{B}(x,y)\|\boldsymbol{A}(x,y)\}, \tag{8}$$

其中，“||”表示拼接操作，这种编码方式同时包含了目标概率的空间分布、无人机的高度信息和协同观测信息，为主动搜索的决策过程提供了全面的输入。

2) 动作空间：本文的主动搜索场景中，三维空间中单个无人机的动作空间A为考虑沿3个轴（X、Y和Z）中任意一个轴的单位距离位移。A的每个元素都是一个三元组，定义了三维空间中的独特运动，总共包含27种可能的动作。

3) 奖励构造：本文的主动搜索框架中，奖励整形旨在指导无人机进行高效的探索和目标检测。奖励函数由两个主要组件组成，分别为熵减少奖励和目标与空白空间识别奖励。

熵减少奖励是基于置信矩阵熵的减少。对于每个网格单元，熵计算为$I_t^p=-\phi_t^p\ln(\phi_t^p)$，其中，$\phi_t^p$表示网格单元在时间$t$的置信度分数。然后将从时间$t$到$t$+1的总熵减少量计算为所有网格单元的熵之差

$$r_1=\Gamma_{t+1}-\Gamma_t \tag{9}$$

其中，Γ_t是置信矩阵在时间t的总熵，即所有网格单元的熵之和

$$\Gamma_t=-\sum_{p\in M}\phi_t^p\ln(\phi_t^p) \tag{10}$$

目标与空白空间识别奖励旨在区分目标位置和空白空间的置信度更新。如果目标网格的置信度增加或者对空白网格的置信度降低，则将在无人机观察和置信度地图更新后分配奖励。目标和空白单元格的奖励设置为相同的值以平衡检测目标和清除空白空间的重要性

$$r_2=w_{\text{h}}\cdot\sum_{p_{\text{h}}\epsilon M_{\text{h}}}\triangle\phi_t^{p_{\text{h}}}+w_{\text{e}}\cdot\sum_{p_{\text{e}}\epsilon M_{\text{e}}}\triangle\phi_t^{p_{\text{e}}} \tag{11}$$

其中，M_{e}和M_{h}分别代表空白网格和目标网格。

最终时间折扣奖励是r_1和r_2的组合，并带有时间折扣因子。时间折扣因子$(T_{\max}-t)/T_{\max}$考虑了时间的紧迫性，鼓励代理在合理的时间范围内实现其目标

$$r=(w_1\cdot r_1+w_2\cdot r_2)\cdot\frac{T_{\max}-t}{T_{\max}} \tag{12}$$

其中，w_1和w_2是可以调整的权重，以平衡熵减少的重要性以及目标和空白空间之间的区分。这种奖励塑造策略确保了代理专注于全面了解环境，同时又专注于完成检测目标的特定任务，在探索和利用之间取得平衡。

4) 强化学习算法：本文的无人机主动搜索模型中，任何给定时间的观测集$\boldsymbol{S}_t=\{\boldsymbol{B}(x,y),\boldsymbol{A}(x,y)\}$都经历关键的归一化和缩放过程，以确保输入数据规模的一致性。置信度矩阵$\boldsymbol{B}$通过减去其均值来标准化。高度图矩阵$\boldsymbol{A}$使用如下方式归一化

$$\boldsymbol{A}'(x,y)=\frac{\boldsymbol{A}(x,y)}{h_{\text{High}}-h_{\text{Low}}} \tag{13}$$

a) CNN-LSTM编码器：本文的无人机主动搜索模型中，利用CNN-LSTM[29-30]架构作为演员—评论家网络的共享编码器。最初，CNN层处理标准化和缩放的观测值$\boldsymbol{S}_t$。这些层从置信度矩阵、高度数据和观测输入中的各种数据类型中熟练地提取空间特征，并将其转换为适合时间分析的统一特征集。提取这种空间特征之后，编码器的LSTM组件开始发挥作用，用于处理观测的时间动态。LSTM的架构专为记忆和顺序数据处理而设计，使其在无人机监视和搜索环境中非常有效。它可以随着时间的推移跟踪相关数据，从而根据当前和过去的信息做出明智的决策。

b) PPO：本文使用PPO算法[31]，利用集中式训练和执行框架，该框架特别适用于本文的主动搜索

任务中多个无人机的协调操作。这个集中式框架确保所有无人机都遵守统一的策略，这对于在复杂的搜索操作中保持团队合作和效率至关重要。该算法采用广义优势估计（GAE, generalized advantage estimation）来计算优势函数，对于评估当前政策$\pi_\theta(a|s)$下的行动绩效至关重要。

CARL方法与PPO算法相结合，能够有效地同步所有无人机的学习和操作，确保它们在主动搜索场景典型的动态和不可预测的环境中协作运行。这种方法不仅提高了搜索策略的整体有效性，还确保了学习策略在所有无人机上的统一应用，从而最大限度地提高了系统的总体性能。

3　实验评估

3.1　实验设置

实验在模拟的三维环境中进行，该环境被离散为80×80×100个网格单元，对应于800 m×800 m×100 m的真实世界搜索空间。这种离散化是根据无人机在不同飞行高度的视场计算确定的。此环境中随机分布了10~50个目标，以模拟真实灾害场景中幸存者的稀疏分布。分别使用2、4、6、8和10架无人机评估模型在搜索任务中的性能。此外，为了保证实验的可靠性，在每种配置下的每次重复试验中均会选择随机的无人机初始位置。6架无人机，50个目标的实验示例如图4所示，图4(a)为三维视图，图中各种颜色的十字和线条分别表示不同的无人机及其轨迹，图4(b)为置信图，该图直观地展示了对目标位置的当前估计，颜色越深表示概率越高。从图4中可以看出无人机在搜索空间中分散目标方面的动态协作，此时处于初期探索阶段，因此无人机选择相互配合优先提高整体覆盖率。实验参数配置见表1。

3.2　基线与指标

实验中，对本文所提方法和3种替代方法进行了比较，以验证所提方法的有效性。

1) 噪声感知的汤普森采样（NATS, noise-aware Thompson sampling）[9]方法将并行异步Thompson采样用于去中心化多智能体算法，平衡了探索和利用。它结合了实际环境下传感器考虑的因素，包括物体检测的不确定性随距离的增加和传感器视场的限制。

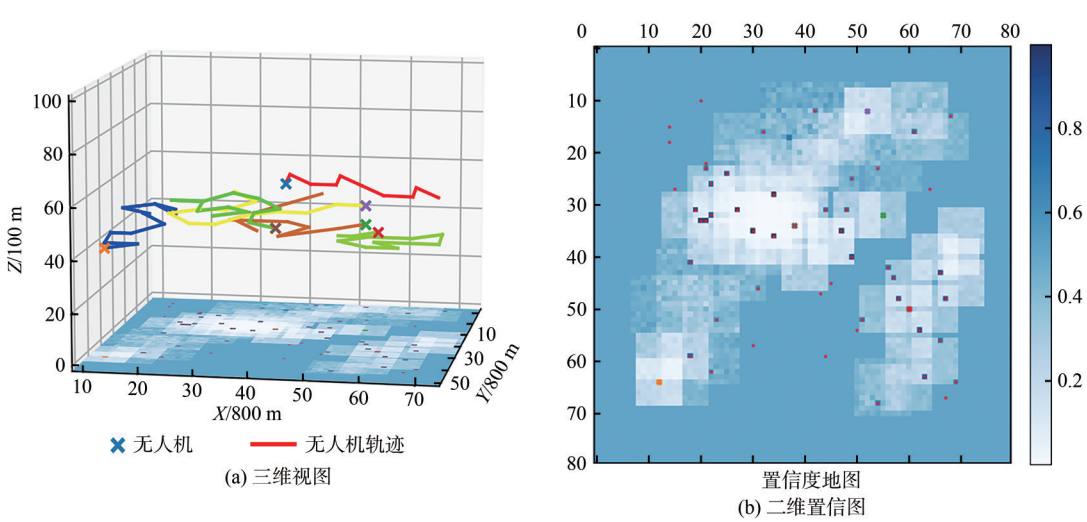
2) 信息增益（IG, information gain）方法采用式(4)中描述的方法来计算信息增益，并利用贪婪算法来制定无人机下一步决策的策略。

3) 随机策略（Random based）为无人机通过随机游走策略探索环境并定位目标。

本文所提算法的目标是在完全未知的环境中定位目标并确定目标在有限时间内的位置，同时减少环境的不确定性。因此，我们采用目标的完全搜索率来评估算法的性能，其定义为

$$\kappa = \frac{\sum_p 1\{P(\varphi_p) > 0.8, \eta_p = 1\}}{m} \tag{14}$$

此外，本文在实验中评估了不同模型在不同环境中性能的稳定性。在各种环境中进行了50次重复实验，测量了50轮完全搜索率的方差，作为对模型稳定性的评估。



(a) 三维视图　　(b) 二维置信图

图4　6架无人机，50个目标的实验示例

表 1　　实验参数配置

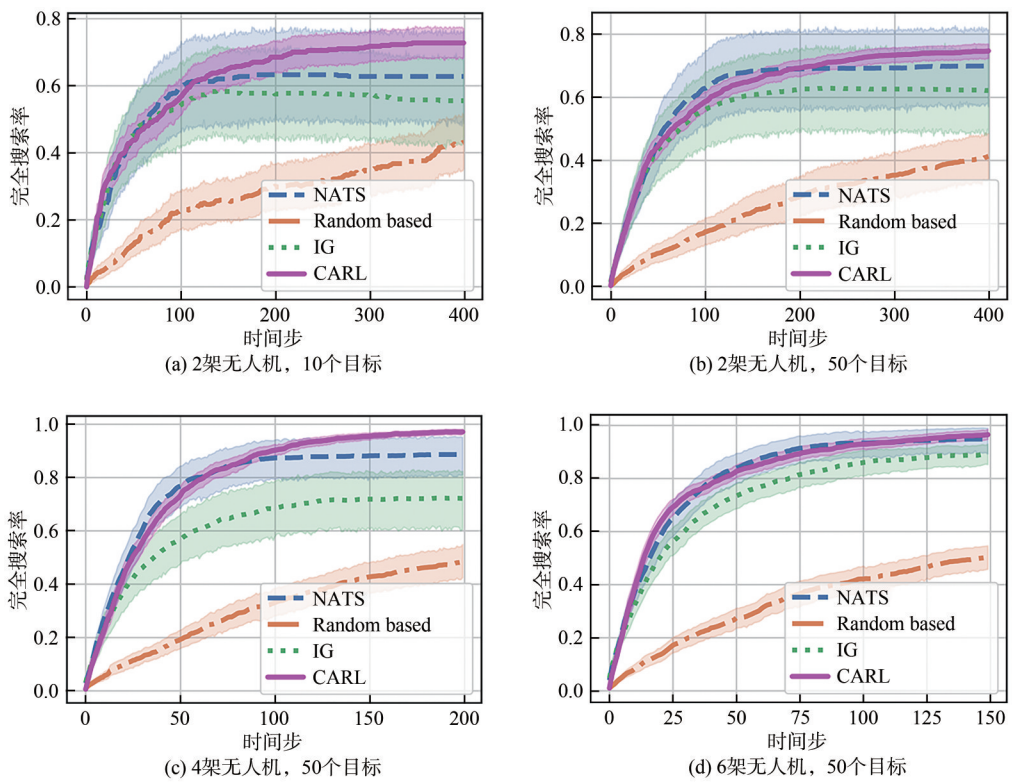
参数名称	参数值
二维网格数 $L_1 \times L_2$	80
目标数量	10~50
高度网格数 L_3	100
无人机高度下限/m	10
无人机高度上限/m	100
最大视场 FOV	10
熵减少奖励权重 w_1	0.1
目标与空白空间识别奖励权重 w_2	0.9
高度探测方差 σ^2/h	h/2 000
最大搜索时间 T_{max}/步	400

3.3　总体性能

1) 完全搜索率：算法完全搜索率随时间的变化如图 5 所示。本文将 CARL 方法与 NATS 方法和 IG 方法在各种情况下进行了比较。实线表示每种方法达到的平均完全搜索率，而阴影区域表示在随机生成的地图上进行的 40 项独立试验的方差。CARL 方法在所有测试配置中始终优于基线，展示了其稳健性和可扩展性。

2) 模型稳定性：算法性能方差随时间的变化如图 6 所示，图 6 显示了在学习阶段 CARL 方法与基线方法相比的性能差异。结果表明，CARL 方法的稳定性明显优于其他基线。其中，初始阶段的波动是正常的，这反映了强化学习的探索过程。随着无人机适应环境，模型性能的波动会减小，这表明策略更加自信和稳定。在模拟结束时，CARL 方法的方差减少，更高、更稳定的完全搜索率表明它已经学习出了一致且有效的搜索策略。此外，当无人机数量增加到 6 个时，随机策略以外的所有算法的性能方差在一定时间步后都开始下降，这表明在较多无人机数量的情况下，各算法均可捕捉所有目标。

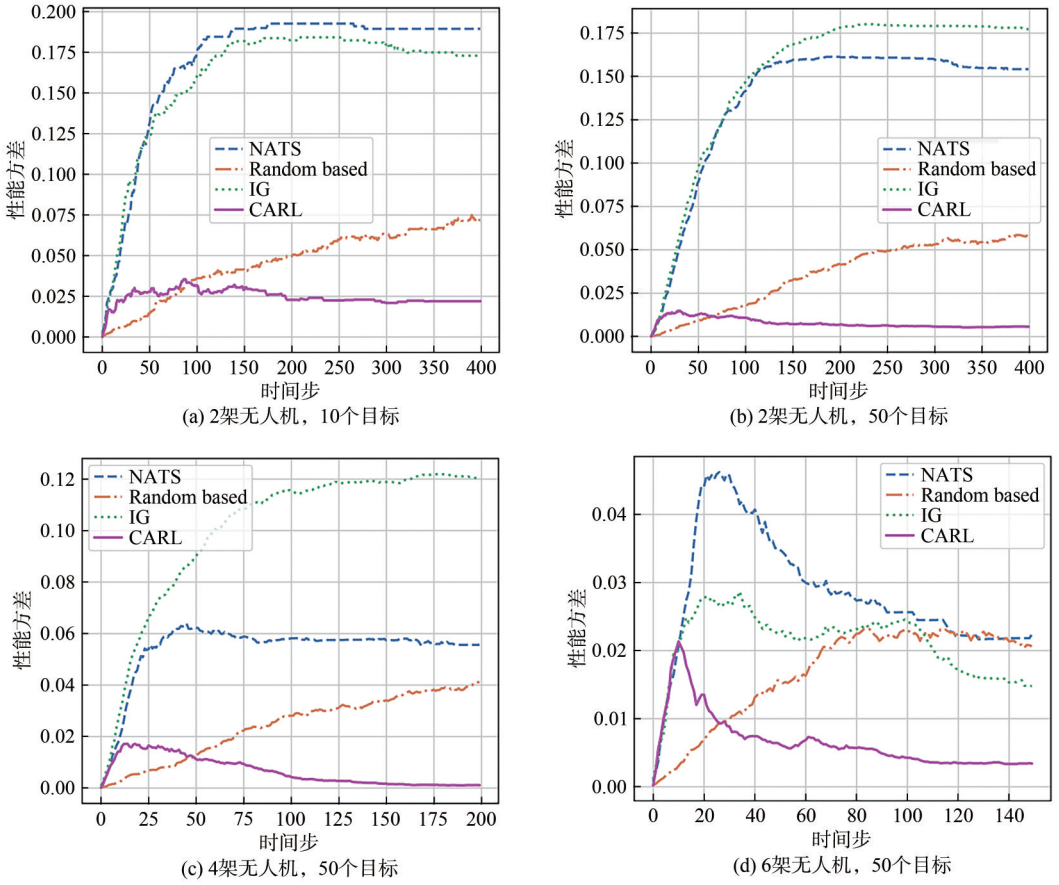
3) 系统鲁棒性：系统鲁棒性的消融实验结果如图 7 所示。该实验突出了本文所提方法在不同操作场景中的性能。图 7(a)验证了系统在不同目标数量情况下的稳定性。结果表明，所提方法保持一致的优越性，证明了其对不同数量的目标具有稳健的适应性。图 7(b)展示了系统性能随无人机数量增加的情况，任务固定为定位相同数量的目标，在这种设置中，随着所有方法增加更多的无人机，完全搜索率都得到了提高，但是本文方法也显示了显著的优越性，尤其在无人机数量更多时，这凸显了本文的多无人机协作策略在实现高效搜索方面的优势。



(a) 2架无人机，10个目标　(b) 2架无人机，50个目标

(c) 4架无人机，50个目标　(d) 6架无人机，50个目标

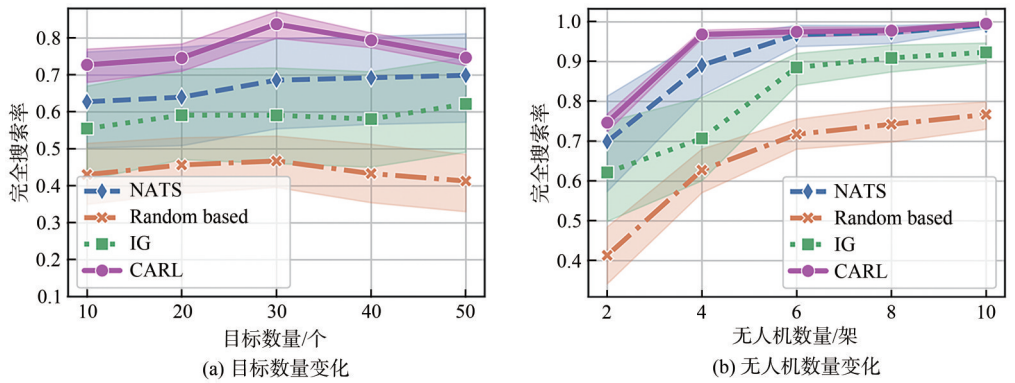
图 5　算法完全搜索率随时间的变化



(a) 2架无人机，10个目标　(b) 2架无人机，50个目标

(c) 4架无人机，50个目标　(d) 6架无人机，50个目标

图6　算法性能方差随时间的变化



(a) 目标数量变化　(b) 无人机数量变化

图7　系统鲁棒性的消融实验结果

4　结束语

本文介绍了CARL方法，这是一种协作式高度自适应强化学习方法，旨在增强无人机集群在主动搜索任务中的执行效能。在不同条件下进行的大量模拟实验表明，CARL方法相较于基线方法具有显著的高效性和稳定性，有望提高无人机集群执行主动搜索任务的效能。未来，将进一步完善CARL方法的架构，考虑更多影响传感器性能的因素，如无人机的朝向以及目标所处的环境因素等，并将CARL方法部署到真实的无人机系统中。

参考文献：

[1] HAKAMI A, KUMAR A, SHIM S J, et al. Application of soft systems methodology in solving disaster emergency logistics problems[J]. International Journal of Industrial and Manufacturing Engineering, 2013, 7(12): 2470-2477.

[2] MEERA A A, POPOVIĆ M, MILLANE A, et al. Obstacle-aware

adaptive informative path planning for UAV-based target search[C]// Proceedings of the 2019 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2019: 718-724.

[3] LYU M Y, ZHAO Y B, HUANG C, et al. Unmanned aerial vehicles for search and rescue: a survey[J]. Remote Sensing, 2023, 15(13): 3266.

[4] POPOVIĆ M, VIDAL-CALLEJA T, HITZ G, et al. Multi-resolution mapping and informative path planning for UAV-based terrain monitoring[C]//Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press, 2017: 1382-1388.

[5] VISERAS A, GARCIA R. DeepIG: multi-robot information gathering with deep reinforcement learning[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 3059-3066.

[6] WANG H Y, XU J G, ZHAO C Y, et al. TransformLoc: transforming MAVs into mobile localization infrastructures in heterogeneous swarms[J]. arXiv preprint, 2024, arXiv: 2403.08815.

[7] WANG H Y, CHEN X C, CHENG Y H, et al. H-SwarmLoc: efficient scheduling for localization of heterogeneous MAV swarm with deep reinforcement learning[C]//Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems. New York: ACM, 2022: 1148-1154.

[8] CHEN X C, WANG H Y, LI Z X, et al. DeliverSense: efficient delivery drone scheduling for crowdsensing with deep reinforcement learning[C]//Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York: ACM, 2022: 403-408.

[9] REN J Y, XU Y G, LI Z X, et al. Scheduling UAV swarm with attention-based graph reinforcement learning for ground-to-air heterogeneous data communication[C]//Proceedings of the Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing. New York: ACM, 2023: 670-675.

[10] CHEN X C, XIAO Z J, CHENG Y H, et al. SOScheduler: toward proactive and adaptive wildfire suppression via multi-UAV collaborative scheduling[J]. IEEE Internet of Things Journal, 2024, 11(14): 24858-24871.

[11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 779-788.

[12] GHODS R, DURKIN W J, SCHNEIDER J. Multi-agent active search using realistic depth-aware noise model[C]//Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2021: 9101-9108.

[13] GOEL H, LIPSCHITZ L J, AGARWAL S, et al. Reinforcement learning for agile active target sensing with a UAV[J]. arXiv preprint, 2022, arXiv: 2212.08214.

[14] ALAGHA A, SINGH S, MIZOUNI R, et al. Target localization using multi-agent deep reinforcement learning with proximal policy optimization[J]. Future Generation Computer Systems, 2022, 136: 342-357.

[15] MACWAN A, VILELA J, NEJAT G, et al. A multirobot path-planning strategy for autonomous wilderness search and rescue[J]. IEEE Transactions on Cybernetics, 2015, 45(9): 1784-1797.

[16] TOMIC T, SCHMID K, LUTZ P, et al. Toward a fully autonomous UAV: research platform for indoor and outdoor urban search and rescue[J]. IEEE Robotics & Automation Magazine, 2012, 19(3): 46-56.

[17] POPOVIĆ M, HITZ G, NIETO J, et al. Online informative path planning for active classification using UAVs[C]//Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2017: 5753-5758.

[18] ZHENG Y S, CHEN J T. Active search for low-altitude UAV sensing and communication for users at unknown locations[J]. arXiv preprint, 2024, arXiv: 2408.14067.

[19] LI Q Q, TAIPALMAA J, QUERALTA J P, et al. Towards active vision with UAVs in marine search and rescue: analyzing human detection at variable altitudes[C]//Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). Piscataway: IEEE Press, 2020: 65-70.

[20] SHAO R, TAO R T, LIU Y D, et al. UAV cooperative search in dynamic environment based on hybrid-layered APF[J]. EURASIP Journal on Advances in Signal Processing, 2021, 2021(1): 101.

[21] RÜCKIN J, JIN L R, POPOVIĆ M. Adaptive informative path planning using deep reinforcement learning for UAV-based active sensing[C]//Proceedings of the 2022 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2022: 4473-4479.

[22] ZHU H, CHUNG J J, LAWRANCE N R J, et al. Online informative path planning for active information gathering of a 3D surface [C]//Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2021: 1488-1494.

[23] BONO R N, CARPIO R F, GASPARRI A, et al. Information-driven path planning for UAV with limited autonomy in large-scale field monitoring[J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(3): 2450-2460.

[24] ZHU K, HAN B, ZHANG T. Multi-UAV distributed collaborative coverage for target search using heuristic strategy[J]. Guidance, Navigation and Control, 2021, 1(1): 2150002.

[25] IGOE C, GHODS R, SCHNEIDER J. Multi-agent active search: a reinforcement learning approach[J]. IEEE Robotics and Automa-

tion Letters, 2022, 7(2): 754-761.

[26] BANERJEE A, GHODS R, SCHNEIDER J. Cost aware asynchronous multi-agent active search[J]. arXiv preprint, 2022, arXiv: 2210.02259.

[27] ADONI W, LORENZ S, FAREEDH J, et al. Investigation of autonomous multi-UAV systems for target detection in distributed environment: current developments and open challenges[J]. Drones, 2023, 7(4): 263.

[28] JAVAID S, SAEED N, QADIR Z, et al. Communication and control in collaborative UAVs: recent advances and future trends[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(6): 5719-5739.

[29] Rakhlin A. Convolutional neural networks for sentence classification[J]. GitHub, 2016(6): 25.

[30] GRAVES A. Long short-term memory[J]. Supervised Sequence Labelling with Recurrent Neural Networks, 2012: 37-45.

[31] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[J]. arXiv preprint, 2017, arXiv: 1707.06347.

[作者简介]

肖子健（1997– ），男，清华大学深圳国际研究生院硕士生，主要研究方向为多智能体协同、强化学习等。

夏晨钧（1996– ），男，清华大学深圳国际研究生院硕士生，主要研究方向为人机交互、移动传感、强化学习等。

徐杨罡（1999– ），男，清华大学深圳国际研究生院硕士生，主要研究方向为强化学习、大模型智能体、无线通信等。

任纪媛（2000– ），女，清华大学深圳国际研究生院硕士生，主要研究方向为强化学习、移动传感等。

陈鑫磊（1987– ），男，博士，清华大学深圳国际研究生院副教授，主要研究方向为智能物联网、多智能体协同、强化学习、普适计算、脑机接口等。