

面向未知环境机动辐射源的单无人机具身追踪方法

梁宏韬¹, 付杨美子², 万奕尧¹, 刘晓东², 吴启晖^{1*}

(1. 南京航空航天大学电子信息工程学院, 江苏南京 211106; 2. 南昌大学信息工程学院, 江西南昌 330031)

摘要: 在电子侦察、频谱监测与应急搜救等任务中, 单架无人机仅凭接收信号强度指示(Received Signal Strength Indicator, RSSI)对非合作机动辐射源进行被动追踪具有重要的战略与工程意义。然而, 现有RSSI定位方法高度依赖环境传播参数先验, 且将参数估计、状态滤波与路径规划串行割裂, 难以适应以发射功率与路径损耗指数为代表的核心传播参数未知且目标持续机动的动态场景; 同时, 单通道标量RSSI固有的低空间分辨力进一步加剧了探索与利用之间的结构性矛盾。为此, 本文提出一种基于具身智能的单无人机闭环追踪方法, 以结构化不确定度指标作为跨层信息流的统一载体, 将感知、认知与决策通过不确定度的双向传递进行深度耦合, 形成以闭环信息流驱动的协同追踪架构。在感知层, 构建参数先验推断网络, 从历史时空观测序列中提取传播参数的经验分布, 为后续贝叶斯推断提供数据驱动的初始化依据。在认知层, 针对未知传播参数与机动状态的联合估计难题, 设计基于Rao-Blackwellized条件分解的双链解耦推断机制, 外层粒子滤波在线估计大尺度衰落参数, 内层自适应交互多模型滤波器精确推断目标机动状态, 并引入有效样本量监测下的粒子复苏与信念继承机制, 避免参数重采样对状态追踪连续性的破坏。在决策层, 为缓解上述探索-利用矛盾, 构建融合信息增益与模型预测控制的分层规划策略, 并引入可学习的排序门控网络对追踪、探测与保守等候策略进行实时效用评估, 使无人机依据信念不确定度自适应地从广域探索转换为紧密伴飞。仿真结果表明, 所提方法在传播参数完全未知的冷启动条件下各项核心指标均优于所对比的基线方法, 且在无人工标定先验模式下的追踪性能接近乃至在部分指标上超越若干基线在参数已知条件下的理想表现, 验证了所提架构在恶劣感知条件下的自适应能力与鲁棒性。

关键词: 具身智能; 目标导航; 接收信号强度指示; 主动路径规划; 信息驱动决策

基金项目: 国家重点研发计划(No.2023YFB2904500); 长三角科技创新共同体联合攻关(No.2024CSJZN00300)

中图分类号: TN929.5; TN911.7

文献标识码: A

文章编号: 0372-2112(XXXX)XX-0001-15

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20260273

An Embodied Intelligence-Based Method for Single-UAV Tracking of Maneuvering Radiation Sources in Unknown Environments

LIANG Hongtao¹, FU Yangmeizi², WAN Yiyao¹, LIU Xiaodong², WU Qihui^{1*}

(1. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China;

2. School of Information Engineering, Nanchang University, Nanchang, Jiangxi 330031, China)

Abstract: Passive tracking of uncooperative maneuvering radiation sources by a single unmanned aerial vehicle (UAV) using only received signal strength indicator (RSSI) measurements is of considerable strategic and engineering value for missions such as electronic reconnaissance, spectrum monitoring, and emergency search and rescue. However, existing RSSI localization methods rely heavily on environmental propagation priors and treat parameter estimation, state filtering, and path planning as isolated serial processes, leaving them unable to handle dynamic scenarios in which core propagation parameters such as transmit power and path loss exponent are unknown and the target keeps maneuvering. The inherent low spatial resolution of scalar RSSI from a single channel further intensifies the structural conflict between exploration and exploitation. To address these challenges, this paper proposes an embodied intelligence-based closed-loop tracking method for a single UAV, which adopts structured uncertainty metrics as a unified carrier of cross-layer information flow and deeply couples perception, cognition, and decision-making through bidirectional propagation of uncertainty, forming a cooperative tracking architecture driven by closed-loop information flow. At the perception layer, a parameter prior inference network is built to extract the empirical distribution of propagation parameters from historical spatio-temporal observation sequences, providing data-driven initialization for the subsequent Bayesian inference. At the cognition layer, a dual-chain decoupled inference mechanism based on Rao-Blackwellized conditional decomposition is designed for the joint estimation of unknown propagation parameters and maneuvering states, in which the outer particle filter estimates large-scale fading parameters on-

line while the inner adaptive interacting multiple model filter accurately infers maneuvering states of the target. A particle revival mechanism guided by effective sample size monitoring is further introduced together with a belief inheritance mechanism to prevent parameter resampling from disrupting the continuity of state tracking. At the decision layer, to mitigate the exploration-exploitation conflict noted above, a hierarchical planning strategy fusing information gain with model predictive control is constructed, and a learnable ranking gating network is introduced to perform real-time utility evaluation over candidate strategies including tracking, probing, and conservative options, allowing the UAV to adaptively transition from broad-area exploration to tight escort flying according to belief uncertainty. Simulation results show that the proposed method outperforms the compared baseline approaches across all core metrics under cold-start conditions with completely unknown propagation parameters, and the tracking performance under the mode without manual calibration priors approaches and in certain metrics surpasses the ideal performance of several baselines operating with known parameters, which verifies the adaptability and robustness of the proposed architecture under degraded sensing conditions.

Keywords: embodied intelligence; target navigation; RSSI; active path planning; information-driven decision

Foundation Item(s): National Key Research and Development Project (No. 2023YFB2904500); Yangtze River Delta Science and Technology Innovation Community Joint Research (Basic Research) Project (No. 2024CSJZN00300)

0 引言

在现代电子侦察、频谱监测及城市应急搜救等任务中,快速且隐蔽地获取复杂环境下的电磁态势具有至关重要的战略与实际意义^[1-2]。在这些复杂场景中,目标辐射源往往表现为非合作属性,其不仅不会主动广播自身坐标,甚至可能采取机动规避策略以隐藏电磁特征。与此同时,在广域或被破坏的空间内部署传统的固定传感器阵列,往往面临着部署周期长、抗毁坏能力弱以及建设成本高昂等诸多局限^[3]。为了有效打破这种环境约束与目标隐蔽性带来的双重壁垒,依托高机动性的单架无人机执行轻量化的被动侦测任务,成为一种兼顾机动灵活性与工程实用性的技术方案。

在无人机可搭载的各类被动侦测体制中,接收信号强度指示(Received Signal Strength Indicator, RSSI)展现出了极高的硬件经济性与载荷适配度^[4]。文献[3]指出,RSSI无需庞大的方位天线阵列或严苛的时间同步硬件,凭普通射频前端即可实时提取。然而,文献[5]强调由于RSSI本质上仅为反映信号衰减强度的标量,单次采样只能提供高度模糊的距离区间,完全缺失到达角或到达时间差等体制所固有的空间几何约束。更为严峻的是,在真实的无线电传播环境中,单通道RSSI的定位可靠性极度依赖准确的环境先验。文献[6-7]进一步证实,当面临路径损耗指数 n 或阴影衰落方差 σ^2 等核心参数未知的情形时,其测距的Cramér-Rao误差下界显著变差。这种测量维度的先天局限性与动态环境的强不确定性相互交织,使得仅依靠单通道低信息量RSSI序列实现对未知机动辐射源的高精度自主追踪,成为了当前被动定位领域亟待突破的技术瓶颈。

在未知动态环境中实现对机动辐射源的可靠追踪,其核心诉求在于突破对环境先验的强依赖,并将

参数推断、状态估计与主动规划纳入深度的闭环耦合框架。具体而言,文献[8]指出经典RSSI定位方法高度依赖准确的传播参数或指纹库,在参数缺失时极易面临性能失效。尽管文献[6-7]尝试通过在线贝叶斯估计单一参数或多体制联合推断来应对环境时变,且部分研究引入了深度学习网络^[9-11],但文献[9]强调,这些方法均受限于特定训练数据分布或静态拓扑,难以泛化至真实对抗环境。因此,高动态纯RSSI追踪本质上要求进行参数与状态的联合推理。对此,文献[12]论证了Rao-Blackwellized粒子滤波(Rao-Blackwellized Particle Filter, RBPF)能有效规避高维参数空间的粒子退化难题,而文献[13]表明交互多模型(Interacting Multiple Model, IMM)算法可广泛用于应对目标机动模式的高度不确定性。然而,现有前沿工作往往孤立地解决上述某一子问题,尚未在统一框架内同时实现传播参数与机动状态的联合在线推断,也缺乏将推断精度与主动观测规划进行闭环协同的系统级探索。

实现机动辐射源的高精度追踪不仅依赖底层的估计算法,更需要主动路径规划(Informative Path Planning, IPP)获取高质量观测;正如文献[14-15]所强调的,无人机的物理飞行轨迹从根本上决定了观测数据的信息价值。然而,现有规划方法普遍将决策与感知、认知割裂。在现有IPP框架中,文献[16]表明基于Fisher信息的单步贪心策略在面对多模态信念分布时极易陷入局部最优;而文献[17]指出,基于部分可观测马尔可夫决策过程(Partially Observable Markov Decision Process, POMDP)的前瞻规划虽具渐近最优性,但计算开销呈指数级膨胀。尽管文献[18-19]与文献[20]分别引入深度强化学习与连续空间在线算法以缓解计算瓶颈^[21],但这些高阶方法普遍建立在观测模型参数完全已知的严苛假设之上。更为致命的是,上述方法习惯将观测处理、参数与状态联合估

计以及路径规划视为独立的模块进行串行拼接。这种割裂的架构打破了紧密咬合的物理因果环路:决策位置直接决定了感知上限,认知准确度构成了决策的刚性约束,而决策轨迹的空间多样性又反向塑造了参数认知的收敛过程^[21]。这深刻表明,在参数未知的复杂环境下进行机动信号源追踪,必须摒弃简单的模块拼接,转向感知、认知与决策协同优化的系统级设计。上述分析表明,主动感知、信息驱动规划与部分可观测决策框架分别为观测获取、规划优化与序贯推断提供了成熟的理论工具,然而三者在各自经典形态中往往彼此独立运行,尚缺乏在参数未知条件下将感知质量、推断精度与决策收益纳入同一闭环的系统级融合方案。

具身智能的研究范式为上述系统级融合提供了理论支撑。本文将具身智能的核心主张落实为一项可验证的系统组织特征:感知、认知与决策三个功能层通过结构化不确定度指标的双向传递形成不可分割的闭环耦合^[22]。认知层输出的信念熵与有效样本量同时驱动决策层的策略切换与感知层的先验置信度调节,决策层选定的物理动作又直接塑造下一时刻的观测信息价值,由此区别于对上述任一经典框架的单点扩展。该范式克服了脱离物理约束进行孤立数据计算的局限性,强调系统的复杂决策能力应当内生于感知反馈、状态推理与主动执行的持续物理闭环之中^[23]。具体而言,文献[24]深入剖析了智能演化的物理基础,指出持续的感觉运动闭环是系统构建复杂认知能力的核心驱动力,智能体的认知深度与环境的实时物理交互频率息息相关。为确立这一交互机制,文献[25]构建了涵盖动态感知与任务适应的系统级具身架构,论证了在内部世界模型推演与外部主动控制策略之间,必须建立一条高频且稳健的认知与控制反馈回路。在此基础上,文献[26]进一步界定了当前具身智能落地的核心壁垒,即如何跨越计算与物理的异构边界,实现多维感知流、动态认知图谱与物理空间执行动作的深度纠缠与无缝耦合。

针对上述挑战,本文聚焦研究传播参数未知条件下机动辐射源的单无人机具身追踪问题。与传统串行模块化设计不同,本文将具身智能范式落实为感知、认知与决策通过不确定度双向传递形成的闭环耦合架构,使三者在一框架内协同运行。该架构在任务层面呈现三项可辨识的具身特征:不确定度指标作为跨层信息流的统一载体实现双向驱动,物理飞行动作与认知更新之间形成因果闭合的反馈环路,系统的阶段性行为转变由闭环信息流的内在动力学自发涌现而非外部阈值硬编码。具体而言,感知层构建参数先验推断网络以从历史时空序列中提取环境参数的

经验分布;认知层建立双链解耦推断机制以实现传播参数与目标机动状态的联合估计;决策层设计分层规划与排序门控元策略以自适应平衡信息探索与状态利用。三者通过高频物理交互形成动态闭环,在完全缺乏环境先验的条件下实现对机动辐射源的持续追踪。主要贡献总结如下:

(1)针对参数完全缺失与纯标量 RSSI 低空间分辨特征的难题,提出一种基于具身智能理念的闭环系统架构。该架构将参数与状态的联合推断、推断与路径规划的协同优化统一纳入同一物理闭环。具体而言,感知层从历史 RSSI 序列中估计传播参数的先验分布;认知层通过双链解耦架构并行估计传播参数与目标机动状态;决策层依据信念不确定度在追踪、探测与保守策略间动态切换,引导无人机主动飞向高信息增益区域。该紧密嵌套的系统级设计打通了探测反馈、状态更新与轨迹执行的高频交互闭环,为不确定环境下的持续伴飞提供了系统级架构方案。

(2)面向未知时变传播参数与机动目标状态的强耦合难题,建立了一套基于双链解耦与分层元策略门控的联合推断与主动规划方法。在推断阶段,采用 Rao-Blackwellized 粒子滤波在线估计大尺度衰落参数,并引入基于有效样本量监测的粒子复苏机制以克服高维参数空间退化问题,嵌套于其中的状态滤波链则利用自适应交互多模型滤波器,随目标运动模式动态调节状态转移概率。在规划阶段,构建了融合多步信息增益探测与模型预测控制的层级化框架,并引入可学习的排序网络对追踪、探测和保守等候选动作策略进行实时打分,实现了在不同信念不确定度下状态利用与信息探索的自适应平衡。

(3)仿真结果表明,所提方法在包含多种复杂机动轨迹的测试中,各项核心指标均优于所对比的独立滤波贪心策略与在线规划算法。在传播参数完全未知的冷启动条件下,本文方法的追踪性能接近乃至在部分指标上超越若干基线在真实参数已知条件下的表现,初步验证了所提架构在恶劣感知条件下的自适应能力。

1 系统模型与问题建模

1.1 系统架构模型

考虑如图 1 所示面向未知环境机动辐射源的单无人机追踪系统。系统由一架配备多频段被动射频前端的无人机与一个非合作机动辐射源构成,二者均部署于二维离散栅格空间中运行。辐射源不会主动广播自身坐标与运动意图,其发射功率与环境路径损耗指数等核心传播参数均为先验未知。无人机在飞行过程中通过 RSSI 扫描动态采集辐射源的电磁信

号,提取多频段 RSSI 标量序列作为系统唯一的外部观测输入,并在感知、认知与决策的持续闭环交互中在线更新对环境参数与目标状态的联合信念,据此自主生成下一时刻的导航指令。

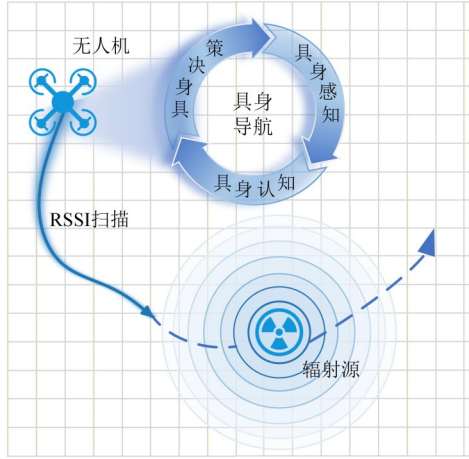


图1 系统模型

Figure 1 System model

系统运行于离散时间步模式,单步时间跨度内无人机的空间位移与观测采样同步完成。该离散步长的选取需兼顾无人机物理机动惯性与信号采样的空间相关性,以确保相邻时间步间系统状态的平稳过渡。在每个时间步的运行周期内,无人机首先基于当前空间位姿获取实时电磁观测 RSSI 序列,随后更新包含未知环境参数与目标运动状态的内部综合信念,最后依据该信念自主生成下一时刻的飞行引导指令。

1.2 实体运动学模型

基于二维离散栅格空间的设定,假设该空间的长宽尺寸为 $H \times W$ 且总栅格数为 HW 。无人机作为主动执行体,其空间状态由二维位置向量 $\mathbf{p}_t^{\text{uav}} = (x_t^u, y_t^u)$ 与离散化朝向 $\phi_t^u \in \{0, 1, \dots, 7\}$ 联合确定,记为 $\mathbf{s}_t^{\text{uav}} = (\mathbf{p}_t^{\text{uav}}, \phi_t^u)$ 。受限于平台物理机动惯性,其控制指令被映射为前进、左转前进与右转前进三种基本动作,即 $a_t \in A = \{0, 1, 2\}$ 。令 $\Delta\phi(a_t)$ 表示具体动作对应的朝向索引增量,无人机的运动学转移方程定义为

$$\phi_{t+1}^u = (\phi_t^u + \Delta\phi(a_t)) \bmod 8 \quad (1)$$

$$\mathbf{p}_{t+1}^{\text{uav}} = \mathbf{p}_t^{\text{uav}} + \Delta_s \cdot \text{dir}(\phi_{t+1}^u) \quad (2)$$

其中, Δ_s 为单步固定的飞行步长; $\text{dir}(\cdot)$ 为将离散朝向映射为二维方向向量的函数。当规划轨迹触及环境边界时,系统采用弹性反射机制处理边界约束,即截断越界坐标并翻转垂直于边界的速度分量,以确保无人机始终在有效空域内安全伴飞。同时,定义目标辐射源的空间状态为 $\mathbf{x}_t^{\text{tgt}} = (\mathbf{p}_t^{\text{tgt}}, \psi_t)$, 其中 $\mathbf{p}_t^{\text{tgt}}$ 为其二维物理坐标; ψ_t 为其离散运动朝向。

1.3 信号传播与观测模型

在无线电信号传播过程中,接收功率呈现出高度的非线性衰减特征。设无人机与机动辐射源之间的实际物理距离为 $d(t) = \Delta_g \|\mathbf{p}_t^{\text{uav}} - \mathbf{p}_t^{\text{tgt}}\|_2$, 其中 Δ_g 为栅格尺度因子。

无人机配置了同步工作在四个离散频段上的射频前端,针对第 i 个频率通道 f_i , 其接收功率 $P_{\text{rx},i}(t)$ (单位 dBm) 服从包含小尺度阴影衰落的对数距离路径损耗模型:

$$P_{\text{rx},i}(t) = P_{\text{tx}} - L_0 - 10n \log_{10} \left(\frac{d(t)}{d_0} \right) - 10\alpha_f \log_{10} \left(\frac{f_i}{f_0} \right) + w_i(t) \quad (3)$$

其中, P_{tx} 为辐射源的发射功率; n 为路径损耗指数; α_f 为频率衰减因子; L_0 为参考距离 d_0 处的基准损耗。阴影衰落噪声服从零均值高斯分布 $w_i(t) \sim \mathcal{N}(0, \sigma_{\text{dB}}^2)$ 。系统将环境传播参数封装为核心参数向量 $\boldsymbol{\theta} = [P_{\text{tx}}, n]^T$ 。

令 $\hat{P}_{\text{rx},i}(\mathbf{x}_t^{\text{tgt}}, \mathbf{s}_t^{\text{uav}}, \boldsymbol{\theta})$ 表示基于上述模型计算所得的无噪声理论接收功率。令 $\mathbf{z}_t = [z_{t,1}, z_{t,2}, \dots, z_{t,4}]^T$ 为当前时刻的原始观测向量(其各分量 $z_{t,i}$ 对应第 i 个频段的实际观测功率 $P_{\text{rx},i}(t)$, 单位 dBm), 给定无人机位置、目标状态与确定的传播参数,其联合似然函数可建模为相互独立的高斯概率密度乘积形式:

$$p(\mathbf{z}_t | \mathbf{x}_t^{\text{tgt}}, \mathbf{s}_t^{\text{uav}}, \boldsymbol{\theta}) = \prod_{i=1}^4 \frac{1}{\sqrt{2\pi} \sigma_{\text{dB}}} \exp \left(- \frac{(z_{t,i} - \hat{P}_{\text{rx},i}(\mathbf{x}_t^{\text{tgt}}, \mathbf{s}_t^{\text{uav}}, \boldsymbol{\theta}))^2}{2\sigma_{\text{dB}}^2} \right) \quad (4)$$

当参数 $\boldsymbol{\theta}$ 完全未知时,较高的发射功率与较大的路径损耗指数在特定距离内能产生极其相似的似然值。这种固有的几何耦合模糊性导致基于单点观测的估计极其困难,系统必须依靠在广域空间内积累具备高空间分集的观测序列来消除这一模糊性。

为了消除设备硬件差异并提高后续神经网络的数值稳定性,各通道的原始接收功率将被映射至有界区间,形成归一化观测向量 $\tilde{\mathbf{z}}_t \in [0, 1]^4$ 供系统感知层的神经网络模块处理。

1.4 具身追踪问题建模

基于上述物理运动与观测模型,未知环境中的单机追踪任务可形式化为一个高维部分可观测马尔可夫决策过程。令目标状态为 $\mathbf{x}_t^{\text{tgt}}$ 、机动模式为 m_t 、传播参数为 $\boldsymbol{\theta}$, 在获取历史观测序列 $\mathbf{z}_{1:t}$ 后,系统的联合信念状态定义为目标状态、机动模式与传播参数的联合后验分布:

$$b_t = p(\mathbf{x}_t^{\text{tgt}}, m_t, \boldsymbol{\theta} | \mathbf{z}_{1:t}), \quad (5)$$

后续各节均以 b_t 指代该分布。认知层在每个时间步同步输出结构化不确定度特征 $(\bar{H}(t), \bar{\text{ESS}}_t, \rho_{\text{top2}})$, 分别表征空间模糊度、参数收敛质量与位置信念多模态程度。

依据贝叶斯法则, 该联合后验可分解为参数后验与条件状态后验的乘积:

$$p(\mathbf{x}_t^{\text{tgt}}, m_t, \boldsymbol{\theta} | \mathbf{z}_{1:t}) = p(\boldsymbol{\theta} | \mathbf{z}_{1:t}) \cdot p(\mathbf{x}_t^{\text{tgt}}, m_t | \boldsymbol{\theta}, \mathbf{z}_{1:t}), \quad (6)$$

该分解揭示了问题的内在耦合结构, 参数后验 $p(\boldsymbol{\theta} | \mathbf{z}_{1:t})$ 的收敛质量取决于观测序列 $\mathbf{z}_{1:t}$ 的空间分集特性, 而空间分集由无人机飞行轨迹 $\mathbf{s}_{0:t}^{\text{uav}}$ 直接决定; 条件状态后验 $p(\mathbf{x}_t^{\text{tgt}}, m_t | \boldsymbol{\theta}, \mathbf{z}_{1:t})$ 的推断精度又以参数后验的准确性为前提; 而飞行轨迹的规划质量则依赖于当前信念状态对目标位置与环境参数的估计可靠度。

定义归一化位置信念分布熵

$$\bar{H}(t) = H(b_t^{\text{pos}}) / \log(HW), \text{ 其中 } b_t^{\text{pos}}(x, y) = \sum_{\psi} \sum_m \int b_t d\boldsymbol{\theta}$$

为对朝向、机动模式与参数边缘化后的目标位置信念分布。系统的核心目标是通过寻找最优元策略 π_{gate} , 在有限时间步 T 内最大化期望累积折扣收益:

$$\begin{aligned} \max_{\pi_{\text{gate}}} \mathbb{E}_{\pi_{\text{gate}}, \mathcal{T}, \mathcal{O}} \left[\sum_{t=0}^{T-1} \gamma^t \mathcal{R}(b_t, a_t) \right] \\ \text{s.t. } b_{t+1} = \Phi(b_t, a_t, \mathbf{z}_{t+1}) \\ \mathbf{s}_{t+1}^{\text{uav}} = f_{\text{kinematics}}(\mathbf{s}_t^{\text{uav}}, a_t), \quad a_t \in A \end{aligned} \quad (7)$$

其中, $\gamma \in (0, 1]$ 为时间折扣因子; $\Phi(\cdot)$ 为符合标准贝叶斯滤波法则的信念状态更新算子; $f_{\text{kinematics}}(\cdot)$ 封装了 2.2 节中定义的无人机运动学方程及边界反射约束。即时综合奖励 $\mathcal{R}(b_t, a_t)$ 由多项竞争性指标线性组合而成, 表示为

$$\mathcal{R}(b_t, a_t) = \mathcal{R}_{\text{track}}(t) + \lambda_{\text{info}} \Delta \bar{H}(t) - \lambda_{\text{mm}} \rho_{\text{top2}}(t) - \lambda_{\text{turn}} C_{\text{turn}}(a_t) \quad (8)$$

其中, $\mathcal{R}_{\text{track}}(t) = 1 - \min(1, d(t)/r_{\text{track}})$ 为稠密追踪收益, $d(t)$ 为仿真器提供的真实物理距离 (仅在训练阶段可获取), 当距离小于预设捕获半径 r_{track} 时系统取得极值收益; λ_{info} 、 λ_{mm} 与 λ_{turn} 均为非负权重系数。 $\Delta \bar{H}(t) = \bar{H}(t-1) - \bar{H}(t)$ 为信息增益, 衡量执行动作前后位置信念熵的下降量, 激励系统主动进行空间探索以消除认知不确定性。 $\rho_{\text{top2}}(t)$ 为信念集中度惩罚项, 定义为位置信念分布中次优概率峰值与全局最优概率峰值的比值, 惩罚因观测分集不足导致的信念分布向多个虚假峰值发散, 迫使推断系统收敛至唯一真实解。 $C_{\text{turn}}(a_t) = I(|\Delta \phi(a_t)| > 0)$ 为转弯惩罚项, 其中 $I(\cdot)$ 为示性函数, 抑制无人机的高频无效盘旋以保障飞行轨迹的平滑性与能量效率。

上述问题建模的成立依赖于四项基本前提: 目标

运动服从离散时间一阶马尔可夫过程, 给定当前状态与机动模式后转移概率与历史轨迹无关; 给定收发位置与传播参数后各频段观测相互独立且噪声独立同分布; 传播参数向量 $\boldsymbol{\theta}$ 在单次追踪周期内保持准静态; 无人机通过累积具备充分空间分集的观测序列能够唯一辨识传播参数。当目标运动呈现长程时序相关、跨频段存在相关衰落、传播参数发生快速跳变或无人机活动范围过于狭窄时, 上述前提将分别被削弱, 方法的适用性需据此重新评估。

然而, 即使上述前提得以满足, 求解该问题仍是在高维混合状态空间中进行序贯决策, 面临计算与控制的双重挑战。一方面, 单步 RSSI 信息量极低, 连续的传播参数与离散机动模式深度耦合, 直接粒子化求解极易导致权值退化, 系统必须引入参数经验先验并设计高效的条件分解机制, 才能在有限算力下实现并行在线收敛。另一方面, 前述“参数辨识依赖空间分集”的前提引发了深度的探索-利用困境: 逼近目标虽能获取即时追踪收益, 却因观测几何单一而不利于参数辨识; 大范围机动虽能丰富观测分集以加速参数收敛, 却会牺牲追踪紧密度。因此, 如何在感知受限且缺乏环境先验的条件下, 同时突破联合推断的维数瓶颈与探索利用的结构性矛盾, 构成了求解该追踪问题的核心关键。

2 面向未知环境机动辐射源的单无人机具身追踪方法

在参数完全未知的环境下, 无人机的物理飞行轨迹直接决定观测数据的空间分集与有效信息量, 而观测数据的质量构成了状态推断精度的上限, 推断精度又反过来约束决策质量。这种感知、认知与决策之间的强物理耦合关系要求系统必须在统一框架内实现三者的协同闭环, 而非串行拼接独立模块。

为此, 本文提出一种感知、认知与决策深度耦合的单无人机具身追踪算法, 其整体运行架构如图 2 所示。系统划分为具身感知、具身认知与具身决策三个核心层级。具身感知层通过时间滑窗机制聚合历史多通道射频观测与无人机空间位姿, 依托深度网络映射为传播参数的经验先验分布与空间约束信息, 为认知层提供数据驱动的初始化依据。具身认知层将目标空间坐标、机动模式与未知衰落参数的联合后验形式化为统一的高维信念状态, 通过解耦推断机制在有限算力下实现联合收敛, 同时向决策层输出量化的不确定度量。具身决策层采用分层规划策略, 底层快速生成局部候选动作集, 上层依据信念不确定度对候选动作进行效用评估与动态排序, 最终选定的执行动作改变无人机的空间观测位置, 驱动下一时刻高价值

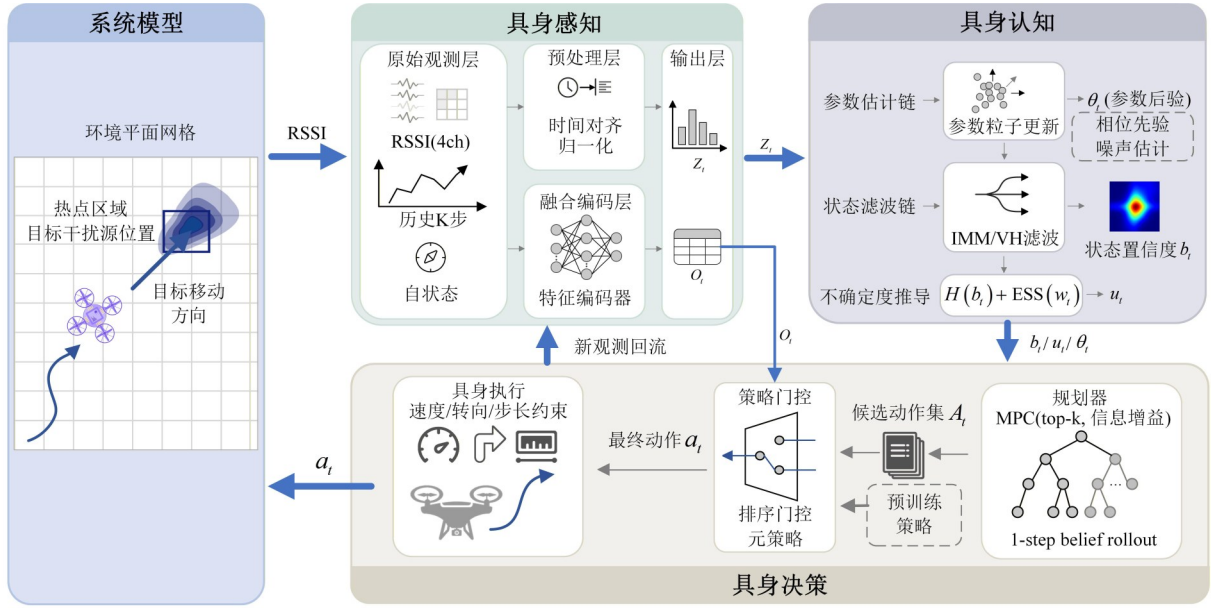


图2 单无人机组身追踪算法整体架构

Figure 2 Overall architecture of the proposed single-UAV embodied tracking method

观测数据的采集,从而闭合感知反馈、认知更新与规划执行的高频物理交互环路。

2.1 具身感知层:时空观测特征提取与编码

为克服单步观测稀疏的物理局限,感知层在每个离散时间步对原始观测执行归一化与坐标系变换后,同步构建固定长度为 L 的联合时空序列 $\mathbf{H}_t \in \mathbb{R}^{L \times 6}$ (各行依次包含无人机相对位移 $\Delta x_t, \Delta y_t$ 与四通道归一化RSSI $\tilde{z}_{t,1}, \tilde{z}_{t,2}, \dots, \tilde{z}_{t,4}$)与有效性掩码 $\mathbf{v}_t \in [0, 1]^L$ 。此时序聚合机制将标量信号波动与无人机物理轨迹进行几何耦合,直接从历史序列中提取环境参数先验与空间约束,为认知层推断提供特征基础。

2.1.1 基于时序特征的环境参数先验推断

针对环境参数缺失易引发贝叶斯滤波器初期发散的问题,感知层构建了参数先验推断网络。该网络直接从联合时空序列 \mathbf{H}_t 中提取空间信号衰落特征,以概率分布的形式回归出未知传播参数向量 θ 的经验先验均值 μ_θ 与标准差 σ_θ 。此数据驱动机制弥补了物理先验的空白,为高维联合推断提供初始化依据。

为保障先验信息的数值鲁棒性,网络输出端引入了多级约束机制。系统在将参数均值 μ_θ 严格截断于物理可行区间 $[\theta_{\text{low}}, \theta_{\text{high}}]$ 的同时,利用掩码计算有效观测步数 $|\mathbf{v}_t|$ 以构建不确定度自适应保护策略。有效步数越少,系统动态设定的标准差 σ_θ 下限越高,从而在观测样本稀疏阶段强制网络保持高方差输出以抑制伪收敛;若有效步数低于启动阈值,系统则直接回退至广域高斯分布 $\mathcal{N}(\mu_{\text{mid}}, \text{diag}(\sigma_{\text{fallback}}^2))$ (其中 $\mu_{\text{mid}} = (\theta_{\text{low}} + \theta_{\text{high}})/2$)以确保推断链路的安全接管。

2.1.2 数据驱动的空间偏移预测与高阶输出

作为环境参数推断的辅助增强,感知层平行构建空间偏移预测模块。该模块默认启用并参与每个时间步的标准运行流程,但其对融合信念的贡献幅度由置信度门控权重 α_{off} 自适应调节:当预测不确定性较高时 α_{off} 趋近于零,模块影响自动消退,系统的核心追踪能力完全由认知层的双链解耦推断独立保障。该模块通过时序深度网络直接回归目标方位偏移,并将其转化为二维高斯软约束分布 $\mathcal{G}^{\text{off}}(x, y)$ 。系统依据预测标准差的特征标量 $\bar{\sigma}^{\text{off}}$ 动态解算置信度门控权重 $\alpha_{\text{off}} = \text{clip}(w_{\text{off}} \cdot \exp(-\kappa \cdot \bar{\sigma}^{\text{off}}), 0, 0.95)$ (其中 w_{off} 为基础融合权重系数, κ 为不确定度衰减率),据此与认知层输出的含朝向基准空间信念 $\bar{b}_t(x, y, \psi)$ 实施加权融合,并在融合后重新进行概率归一化:

$$\bar{b}_t^{\text{fused}}(x, y, \psi) \propto \bar{b}_t(x, y, \psi) \cdot [(1 - \alpha_{\text{off}}) + \alpha_{\text{off}} \cdot \mathcal{G}^{\text{off}}(x, y)] \quad (9)$$

上述参数先验与空间偏移预测构成了感知层的高阶特征输出 \mathbf{o}_t 。该高阶表征与原始dBm域观测 \mathbf{z}_t 一同传入后续认知层: \mathbf{o}_t 在冷启动阶段为参数粒子初始化提供知情先验, \mathbf{z}_t 在每个时间步驱动贝叶斯似然更新。

2.2 具身认知层:双链解耦联合推断

为突破连续参数与离散机动模式深度耦合带来的维数灾难,认知层建立了一套基于 Rao-Blackwellized 定理的双链解耦推断机制,通过粒子集与嵌套滤波器的深度协同,在有限算力下实现了环境参数与目标机动状态的联合收敛。

2.2.1 双链联合推断总体框架

双链推断框架的核心思想是利用条件独立性对高维联合后验概率密度进行严格的降维分解。令目标状态为 $\mathbf{x}_t^{\text{tgt}}$ 、机动模式为 m_t 、传播参数为 θ ，在获取当前时刻 t 的历史观测序列 $\mathbf{z}_{1:t}$ 后，其联合后验分布被贝叶斯法则解析分解为边缘参数后验与条件状态后验的乘积：

$$p(\mathbf{x}_t^{\text{tgt}}, m_t, \theta | \mathbf{z}_{1:t}) = p(\theta | \mathbf{z}_{1:t}) \cdot p(\mathbf{x}_t^{\text{tgt}}, m_t | \theta, \mathbf{z}_{1:t}) \quad (10)$$

上述恒等分解能够转化为可计算的双链结构，依赖于以下条件独立性前提：给定传播参数 θ 后，观测似然 $p(\mathbf{z}_t | \mathbf{x}_t^{\text{tgt}}, s_t^{\text{nav}}, \theta)$ 被完全确定。基于此，外层采用粒子滤波在低维连续空间 $\theta \in \mathbb{R}^2$ 上近似参数后验分布 $p(\theta | \mathbf{z}_{1:t})$ ，维护规模为 K 的参数粒子集 $\{\theta^{(k)}, w_t^{(k)}\}_{k=1}^K$ ；内层则针对每个参数假设实例化独立的交互多模型 (IMM) 滤波器 $\mathcal{F}^{(k)}$ ，在离散网格空间 $H \times W \times 8$ 上精确推断条件状态分布 $p(\mathbf{x}_t^{\text{tgt}}, m_t | \theta^{(k)}, \mathbf{z}_{1:t})$ 。这种 Rao-Blackwellized 处理通过解析积分精确消除了状态维度方差，使联合估计的总方差相较于在高维空间直接粒子化求解显著降低。两条链路通过条件边缘似然这一唯一接口进行信息交换以更新外层权重，各粒子的状态链推断彼此独立互不干扰。该架构实现了感知与动作的协同赋能，精确复现了传播参数决定衰减物理根因、目标状态决定收发观测几何的因果关系。

2.2.2 参数估计链的知情初始化与权重更新

为缓解参数粒子在缺乏先验时出现的发散问题，参数估计链引入了知情初始化策略。系统提取感知层传导的高阶表征 \mathbf{o}_t 中的先验均值 μ_θ 与标准差 σ_θ ，按预设比例使主体粒子从高斯分布 $\mathcal{N}(\mu_\theta, \text{diag}(\sigma_\theta^2))$ 中采样生成，以利用数据驱动的衰落规律；剩余粒子则从参数可行区间 $[\theta_{\text{low}}, \theta_{\text{high}}]$ 上进行均匀采样，以保障全域空间的探索多样性。

在序贯滤波过程中，参数粒子的权重依据增量观测的预测似然在对数域内累加更新。在时刻 t ，系统提取每个粒子内部 IMM 滤波器的预测似然，即对目标状态与模式积分后的条件边缘似然，表示为

$$p(\mathbf{z}_t | \theta^{(k)}, \mathbf{z}_{1:t-1}) = \sum_m \omega_{t-1}^{m(k)} \int p(\mathbf{z}_t | \mathbf{x}_t^{\text{tgt}}, \theta^{(k)}) p(\mathbf{x}_t^{\text{tgt}} | m, \theta^{(k)}, \mathbf{z}_{1:t-1}) d\mathbf{x}_t^{\text{tgt}} \quad (11)$$

执行权重累加 $\log w_t^{(k)} \leftarrow \log w_{t-1}^{(k)} + \log p(\mathbf{z}_t | \theta^{(k)}, \mathbf{z}_{1:t-1})$ 并归一化。为抑制权重坍塌，系统对归一化权重施加向均匀分布的线性插值正则化。

2.2.3 基于有效样本量的粒子复苏机制

在高维连续参数空间中，粒子滤波固有的权值退化现象易导致搜索陷入局部最优。系统在每个时间步计算有效

样本量 $\text{ESS}_t = 1 / \sum_{k=1}^K (w_t^{(k)})^2$ 及其归一化形式 $\overline{\text{ESS}}_t =$

ESS_t / K 。当 $\overline{\text{ESS}}_t$ 低于安全阈值 τ_{ess} 时，系统强制触发粒子复苏机制。系统按权重升序剔除表现最差的末端粒子，并以当前全体粒子的加权均值 $\bar{\theta}_t$ 与加权标准差为基准，通过高斯抖动重采样生成新参数向量。

信念继承在理论上以最优粒子条件下的空间后验替代了新参数假设下应从头推演的条件后验，引入了一项参数条件偏移近似。该近似的合理性在于：复苏机制以粒子集加权均值为基准生成新参数，使新旧参数在参数空间中的距离有界，对应条件似然场的差异有限；同时新粒子在后续时间步将立即以新参数假设执行标准贝叶斯网格更新，继承带来的初始偏差随新观测的注入被逐步修正，条件后验渐进回归至与真实参数一致的分布。在双链解耦结构下，信念继承仅改变新粒子内层状态链的初始条件，不修改外层参数链的权重更新规则，两条链路的推断一致性在各自的更新逻辑下独立维持。

复苏机制的核心设计在于参数空间与状态空间的严格解耦。在状态维度上，新生成的粒子将直接克隆系统内当前最优粒子的完整 IMM 空间信念状态 $\{b_t^{m(k)}\}_{m=1}^M$ 作为自身的初始网格分布。若对新粒子重新初始化空间信念，则在参数重置的瞬间系统将同步丧失已累积的目标方位估计，导致追踪在复苏后出现短暂但致命的定位断裂，此时无人机需要额外数十步才能重新锁定目标。信念继承机制从根本上保证了参数空间的重新探索不会干扰状态空间的连续追踪，使两条推断链路的演化保持严格的物理独立性。

2.2.4 状态滤波链的自适应交互多模型推断

嵌套于每个参数粒子内部的状态滤波链负责处理目标高机动性带来的转移不确定性。系统依据目标可能的机动特征，在基础定向运动、扩展随机机动与相位感知周期运动（如锯齿形交替转弯与矩形周期直行）三个层级上构建了 M 个并行子滤波器，各子滤波器在离散网格 $H \times W \times 8$ 上维护独立的空间信念分布 $b_t^{m(k)}(x, y, \psi)$ 。在每个离散时间步，子滤波器首先依据对应运动模式的转移概率执行空间预测，随后利用参数假设 $\theta^{(k)}$ 计算各网格点处的理论接收功率与实际观测 z_t 之间的似然，完成贝叶斯网格更新。

为提升模型对复杂机动行为的响应敏捷度，内层 IMM 滤波器引入了基于认知不确定度的双重自适应调节机制。滤波器在执行网格预测前，利用上一时刻归一化位置信念熵 $\bar{H}(t-1)$ 动态解算两类自适应参数。其一为模式切换概率，可以表示为

$$\pi_{\text{sw}}(t) = \pi_{\text{min}} + (\pi_{\text{max}} - \pi_{\text{min}}) \cdot \text{clip}(\eta \cdot \bar{H}(t-1), 0, 1) \quad (12)$$

其中， η 为控制敏感度的缩放因子。当 $\bar{H}(t-1)$ 升高即

目标位置模糊时, $\pi_{sw}(t)$ 增大使得模式转移概率矩阵的非对角元素上升, 促使滤波器在不同机动假设间快速跳转以捕获突发机动。其二为均匀扩散系数 $\epsilon_u^m(t)$, 采用与切换概率类似的线性映射机制, 可以表示为

$$\epsilon_u^m(t) = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) \cdot \text{clip}(\eta \cdot \bar{H}(t-1), 0, 1) \quad (13)$$

通过该式控制各子滤波器中均匀分布的混合比例, 当信念熵较高时增大该系数, 强制概率质量向邻近网格大范围扩散, 防止空间信念在单一区域过度集中。

上述两项自适应调节机制属于对标准 IMM 固定参数设定的启发式扩展, 其理论动机在于建立信念不确定度与滤波器内部参数之间的反馈通道。模式切换概率随信念熵升高而增大, 使滤波器在检测到预测失配时加速模式捕获以增强对突发机动的响应敏捷度; 均匀扩散系数随信念熵升高而增大, 通过动态扩展转移核的支撑集保证先验分布对真实目标位置的覆盖概率, 避免滤波器因先验支撑集不足而陷入局部收缩。两者的合理性基于归一化位置信念熵作为模型预测失配程度间接指示量的经验对应关系。

2.2.5 全局信念融合与不确定度量

在完成双链的并行演化后, 认知模块将分布于各个粒子内部的碎片化后验进行融合。针对目标的空间位置, 系统对所有参数粒子的 IMM 融合信念进行加权求和, 提取出全局空间信念分布:

$$\bar{b}_i(x, y, \psi) = \sum_{k=1}^K w_i^{(k)} \sum_{m=1}^M \omega_i^{m(k)} b_i^{m(k)}(x, y, \psi) \quad (14)$$

其中, $\omega_i^{m(k)}$ 为粒子 k 内第 m 个运动模式的后验概率权重。同时, 系统通过加权求和获取参数后验均值:

$$\hat{\theta}_i = \sum_k w_i^{(k)} \theta^{(k)} \quad (15)$$

以及各机动模式的全局后验概率权重:

$$\bar{\omega}_i^m = \sum_k w_i^{(k)} \omega_i^{m(k)} \quad (16)$$

除一阶统计估计量外, 认知模块还输出一组结构化的不确定度指标: 归一化位置信念熵 $\bar{H}(t) = H(b_i^{\text{pos}}) / \log(HW)$ 表征空间模糊度, 次主峰概率比 ρ_{top2} 表征位置信念的多模态程度, 归一化有效样本量 $\bar{\text{ESS}}$, 反馈参数收敛质量。上述估计结果与不确定度量被统一封装为结构化的综合信念状态, 构成系统对未知物理环境的完整后验表征。

2.3 具身决策层: 分层主动规划与元策略门控

为规避庞大连续空间中的在线前瞻搜索计算瓶颈, 决策层建立了一套分层主动规划与元策略门控机制。底层依托多目标预测控制与信息增益度量提取候选动作子集, 顶层则通过神经网络元策略对候选动作进行动态效用评估与排序, 输出最优物理执行

指令。

2.3.1 基于多目标预测控制的候选轨迹生成

在底层轨迹生成阶段, 决策模块采用滚动时域优化框架。系统设定固定前瞻窗口 N_h , 穷举离散动作空间内所有可行的控制序列 $\mathbf{a}_{t:t+N_h-1} \in A^{N_h}$, 并从位置信念 b_t^{pos} 中提取前 J 个 (Top- J) 概率最高的目标位置假设, 结合各运动模式的后验权重 $\bar{\omega}_t^m$ 进行前向轨迹推演。每条候选无人机轨迹 s 的执行代价被定义为预测距离、越界惩罚与转弯惩罚的折现加权:

$$C_s = \sum_{j,m} w_j^{\text{pos}} \bar{\omega}_t^m \sum_{\tau=1}^{N_h} \gamma^\tau [d_\tau^{s,j,m} + \lambda_{\text{out}} \max(0, d_\tau^{s,j,m} - r_{\text{track}})^2] + \lambda_{\text{turn}} \sum_{\tau=1}^{N_h} \gamma^\tau I(|\Delta\phi(a_\tau^s)| > 0) \quad (17)$$

其中, $d_\tau^{s,j,m}$ 为第 j 个位置假设在模式 m 下与无人机的预测空间距离; w_j^{pos} 为该假设的信念概率权重; r_{track} 为预设追踪半径; λ_{out} , λ_{turn} 分别为越界与转弯惩罚系数; γ 为时间折扣因子。

为克服静态信念无法预见行动后认知收益的局限, 规划器引入了单步信念前推机制。针对每个首步候选动作 $a \in \mathcal{A}$, 系统从当前位置信念 b_t^{pos} 中采样若干目标假设位置, 利用参数后验均值 $\hat{\theta}_t$ 为每个假设生成模拟 RSSI 观测, 并据此计算执行该动作后的期望后验位置信念 $\hat{b}_{t+1}^{\text{pos}}(a)$ 。规划器随后以该期望后验信念替代静态信念重新评估轨迹代价, 将物理动作与认知状态演化进行了数学绑定, 确保生成的轨迹具备后验最优性。

2.3.2 驱动空间探索的信息增益度量

从信息论角度, 执行动作 a 后系统获得的严格信息增益定义为 $\mathcal{I}(a) = H(b_t^{\text{pos}}) - \mathbb{E}_{z_{t+1}|a} [H(b_{t+1}^{\text{pos}})]$ 。鉴于其蒙特卡罗采样的精确计算开销无法满足在线决策的实时性需求, 本文采用测量熵 $H_{\text{meas}}(\mathbf{p}_{t+\tau}^a)$ 作为保序近似替代。该度量量化了候选观测位置处预测 RSSI 分布的散度, 值越大表明预测观测越可区分, 与严格信息增益在单调性上保持一致。此近似在似然函数单峰且信噪比适中时较为稳健; 尽管在严重多模态且观测值域重叠时其保序性可能被削弱, 但 3.3.1 节的单步信念前推机制已将动作对信念结构的影响纳入代价评估, 在一定程度上有效缓解了该偏差。

为平滑单步观测带来的随机波动, 系统沿候选航向计算多步折扣信息增益总量:

$$\mathcal{I}(a) = \frac{\sum_{\tau=0}^{N_{\text{probe}}-1} \gamma_{\text{probe}}^\tau H_{\text{meas}}(\mathbf{p}_{t+\tau}^a)}{\sum_{\tau=0}^{N_{\text{probe}}-1} \gamma_{\text{probe}}^\tau} \quad (18)$$

其中, N_{probe} 为探测前瞻步数; γ_{probe} 为探测折扣因子。

计算所得的多步信息增益 $\mathcal{I}(a)$ 作为奖励项,在位置信念熵 $\bar{H}(t)$ 超过预设安全阈值 H_{safe} 时被动态注入代价函数中。具体而言,引入状态指示函数构建增强代价 $C_s^{\text{aug}} = C_s - \lambda_{\text{ent}}(t) \cdot \mathcal{I}(a_0^s)$, 其中 a_0^s 为轨迹 s 的首步动作,探索权重 $\lambda_{\text{ent}}(t) = \lambda_0 \cdot I(\bar{H}(t) > H_{\text{safe}})$, λ_0 为基础探索系数。该度量机制使得无人机能够在认知模糊时主动向不确定度最高的方向机动,加速内部推断系统的全局收敛。

2.3.3 结构化候选动作集的提取机制

面对穷举生成的海量候选轨迹,决策模块基于物理代价与信息增益,降维提取出一个具备明确语义代表性的三策略候选动作集 $A_i^{\text{cand}} = \{a_i^{\text{track}}, a_i^{\text{probe}}, a_i^{\text{safe}}\}$, 其数学定义为

$$\begin{aligned} a_i^{\text{track}} &= \underset{a \in A}{\operatorname{argmin}} \min_{s: a_0^s = a} C_s^{\text{aug}} \\ a_i^{\text{probe}} &= \underset{a \in A}{\operatorname{argmax}} \mathcal{I}(a) \\ a_i^{\text{safe}} &= \underset{a \in A}{\operatorname{argmin}} \mathbb{E}_{s: a_0^s = a} [C_s] \end{aligned} \quad (19)$$

追踪策略 a_i^{track} 直接抵近目标高概率区域,探测策略 a_i^{probe} 驱动系统执行大范围空间观测,保守策略 a_i^{safe} 在追踪与探测间取折中。该三策略结构将连续的探索与利用权衡空间离散化为具有明确物理语义的候选子集,既保留了自适应决策的灵活性,又将元策略的搜索空间从原始动作空间压缩为语义更丰富的策略空间,显著降低了后续网络的学习难度。上述三策略在探索利用权衡轴上分别定位于极端利用、极端探索与中间折中三个结构性锚点。在本文追踪问题中,探索与利用构成决策空间的主导矛盾维度,原始动作空间中任意候选轨迹的首步动作在该轴上的投影必然落入三个锚点的某个邻域,因此三策略子集在行为模式层面构成了对原始动作空间的有效覆盖。

为确保系统在遭遇毁灭性发散时具备强制纠偏能力,候选提取层内嵌了底层强制覆盖规则。当 $\bar{H}(t)$ 连续多个时间步停滞不降、 $\bar{\text{ESS}}$ 跌破极限阈值,或 ρ_{top2} 过高时,提取模块将阻断常规的策略门控流转,强制系统执行探测策略 a_i^{probe} 。此硬逻辑兜底机制有效防范了神经网络在状态空间边界区域的不可控失效。强制覆盖规则与排序门控网络之间为严格的优先级关系:常态下排序网络拥有完整的决策权限,仅当系统检测到信念熵持续停滞、有效样本量跌破极限阈值或多峰性指标超过安全上限等毁灭性发散征兆时,覆盖规则才阻断网络输出通路并无条件执行探测策略,为系统提供不依赖网络泛化能力的确定性安全兜底。

2.3.4 基于多维特征融合的排序门控网络

候选动作集的最终优选由排序门控网络执行。系统构建紧凑表征向量 $\mathbf{f}_i \in \mathbb{R}^{d_f}$, 严密级联了自身运动

状态(朝向、上一步动作)、信念不确定度特征 ($\bar{H}(t)$ 、 $\bar{\text{ESS}}_i$ 、 ρ_{top2})、参数后验 $\hat{\theta}_i$ 、运动模式权重 $\bar{\omega}_i^m$ 以及各候选动作的归一化相对代价与信息增益。该特征向量被输入至多层感知神经网络中,平行输出各个候选策略对应的效用评估标量 $s_i \in \mathbb{R}^{|A_i^{\text{cand}}|}$, 系统选择效用得分最高的分支执行物理动作映射,即 $a_i = A_i^{\text{cand}}[\operatorname{argmax}_j s_{i,j}]$ 。

在网络的优化训练阶段,由于三个候选动作的代价差异是连续的标量而非绝对的分类型正误,系统摒弃了传统的硬标签分类,转而采用成对排序损失以更好地保序。具体而言,对候选动作对集合 $P = \{(i,j) \mid i \neq j, i, j \in \{1, 2, \dots, |A_i^{\text{cand}}|\}\}$, 排序损失定义为

$$\mathcal{L}_{\text{rank}} = \frac{1}{|P|} \sum_{(i,j) \in P} \log \left(1 + \exp \left(-y_{ij} (s_i - s_j) \right) \right) \quad (20)$$

其中, $y_{ij} = \operatorname{sign}(C_j - C_i)$ 为代价比较指示(代价低者应获更高评分)。为强化网络对全局最优动作分布的拟合能力,系统同步引入软交叉熵损失,表示为

$$\begin{aligned} \mathcal{L}_{\text{ce}} &= - \sum_j p_j^{\text{soft}} \log \frac{\exp(s_j)}{\sum_l \exp(s_l)} \\ p_j^{\text{soft}} &= \frac{\exp(-C_j / \tau_{\text{temp}})}{\sum_l \exp(-C_l / \tau_{\text{temp}})} \end{aligned} \quad (21)$$

其中, τ_{temp} 为控制软标签平滑度的温度参数。网络的总训练目标定义为加权组合 $\mathcal{L} = \mathcal{L}_{\text{rank}} + \lambda_{\text{ce}} \mathcal{L}_{\text{ce}}$ 。选定的最优动作 a_i 最终被输入无人机运动学执行接口,完成从感知到执行的单步物理闭环。

排序门控网络以结构化不确定度特征组作为核心输入,其理论动机在于近似实现 POMDP 框架下的信息驱动决策原则:在高不确定度状态下优先选择探测策略以恢复认知质量,在低不确定度状态下切换至追踪策略以最大化即时追踪收益,由此在认知恢复与追踪利用之间维持动态平衡,以保障长时间尺度上的稳定跟踪。

2.4 系统闭环运行流程与阶段性行为分析

上述三层模块通过前向特征传递与反馈不确定度流构成严密的物理闭环。本节提炼单时间步的完整数学迭代管线(如算法1所示),并剖析系统在不同追踪阶段的自发涌现行为。

该闭环的核心特征在于,系统从探索到利用的阶段性转变并非由人工阈值硬编码,而是由三层交互的内在动力学自发涌现。在冷启动阶段,感知层的高方差先验驱动认知层维持广域信念分布,高信念熵 $\bar{H}(t)$ 经特征级联触发决策层偏向探测策略 a_i^{probe} , 引导无人机大尺度机动以累积观测分集;随着参数在似然竞争中逐步收敛, $\bar{H}(t)$ 的下降引发 IMM 切换概率 $\pi_{\text{sw}}(t)$ 与

扩散系数 $c_u^m(t)$ 同步收缩, 排序网络据此自适应切换至追踪策略 a_t^{track} , 平滑转入稳态伴飞; 若目标剧烈机动致使 $\bar{H}(t)$ 骤升, 该数据流则自动反转以重新消除空间模糊。为支撑上述纯在线的自适应演化, 系统对各模块的训练与部署阶段进行了严格划分。具体而言, 感知层先验网络与决策层排序网络在离线阶段分别以真实参数和仿真真实代价为监督信号完成学习, 随后参数即被冻结; 在在线部署阶段, 认知层的双链解耦推断、决策层的轨迹生成及信息增益度量均为实时计算, 排序网络仅执行前向推理, 且在线代价评估完全基于内部信念分布, 彻底摆脱了对仿真器真实距离信息的依赖。

算法 1 单时间步具身追踪闭环

输入: 观测 z_t , 无人机状态 s_t^{uav} , 时空序列 (H_t, v_t) , 粒子集

$\{\theta^{(k)}, w_t^{(k)}, \mathcal{F}^{(k)}\}_{k=1}^K$

输出: 具身动作 a_t

1. 提取经验先验 $(\mu_\theta, \sigma_\theta) \leftarrow (H_t, v_t) // 2.1.1$ 节
2. 回归空间约束 $(\alpha_{\text{off}}, N_{\text{off}}) \leftarrow (H_t) // 2.1.2$ 节
3. $\pi_{\text{sw}}(t) \leftarrow \pi_{\text{min}} + (\pi_{\text{max}} - \pi_{\text{min}}) \cdot \text{clip}(\eta \cdot \bar{H}(t-1), 0, 1)$
4. for $k = 1, 2, \dots, K$ do
5. IMM 网络运动预测 $\mathcal{F}^{(k)}$.predict $(\pi_{\text{sw}}(t), \epsilon_u^m(t))$
6. 似然更新 $\ell^{(k)} \leftarrow \mathcal{F}^{(k)}$.update $(z_t, \theta^{(k)})$
7. 粒子对数权重累加 $\log w_t^{(k)} \leftarrow \log w_{t-1}^{(k)} + \log \ell^{(k)}$
8. end for
9. 归一化 $\{w_t^{(k)}\}$; 计算 $\text{ESS}_t \leftarrow \left(K \sum_k (w_t^{(k)})^2 \right)^{-1}$
10. if $\text{ESS}_t < \tau_{\text{ess}}$ then 粒子复苏
11. 信念融合 $\bar{b}_t \leftarrow \sum_k w_t^{(k)} \sum_m \omega_t^{m(k)} b_t^{m(k)}$
12. 偏移软约束融合 $b_t^{\text{fused}} \leftarrow [(1 - \alpha_{\text{off}}) + \alpha_{\text{off}} \cdot \mathcal{N}^{\text{off}}]$
13. 边缘化获取位置信念 $b_t^{\text{pos}} \leftarrow \sum_{\psi} \bar{b}_t^{\text{fused}}$
14. 计算 $\bar{H}(t), \rho_{\text{top2}}, \hat{\theta}_t, \bar{\omega}_t^m$
15. 时域代价评估 $\{C_s\} \leftarrow (b_t^{\text{pos}}, \bar{\theta}_t, \omega_t^m, N_h) // 2.3.1$ 节
16. 1-step 信念前推 $\hat{b}_{t+1}^{\text{pos}}(a) \leftarrow f_{\text{Rollout}}(b_t^{\text{pos}}, a, \hat{\theta}_t)$
17. 多步信息增益测量 $\mathcal{I}(a) \leftarrow f_{\text{InfoGain}}(b_t^{\text{pos}}, \hat{\theta}_t)$
18. 三策略候选提取 $A_t^{\text{cand}} \leftarrow (\{C_s^{\text{aug}}\}, \{I(a)\}) // 2.3.3$ 节
19. 特征级联 $f_t \leftarrow [s_t^{\text{uav}}; \bar{H}(t); \text{ESS}_t; \rho_{\text{top2}}; \hat{\theta}_t; \bar{\omega}_t^m; \{C_s\}; \{I\}]$
20. 排序门控优选 $a_t \leftarrow (f_t) // 2.3.4$ 节
21. 运动学执行 $s_{t+1}^{\text{uav}} \leftarrow f_{\text{kinematics}}(s_t^{\text{uav}}, a_t)$
22. 采集新观测 z_{t+1} , 步进 $t \leftarrow t + 1$

系统单时间步的计算复杂度由两项主导。认知层内层状态链对 K 个参数粒子各嵌套 M 个子滤波器, 每个子滤波器在 $H \times W \times 8$ 网格上执行预测与更新, 总复杂度为 $O(K \cdot M \cdot HW)$ 。决策层对 $|A|^N$ 条候选轨迹逐一评估 J 个位置假设与 M 个模式下的预测代价, 总复杂度为 $O(|A|^N \cdot N_h \cdot J \cdot M)$ 。在本文配置 ($K =$

$12, M = 3, H = W = 64, |A| = 3, N_h = 6$) 下, 两项有效计算规模均处于 10^5 量级, 可在 GPU 并行加速下于毫秒量级完成, 与实测的单时间步推理耗时均值 11.6 ms 相吻合。决策层复杂度随 N_h 呈指数增长, 构成更长前瞻视野下的主要可扩展性瓶颈。

3 实验与结果分析

为全面验证所提方法的性能, 本文搭建了二维离散栅格仿真平台, 核心参数如表 1 所示。辐射源配备四频段射频前端, 载频分别为 2.4、3.5、5.2、5.8 GHz。为模拟真实环境中传播条件的不确定性, 仿真环境在每个 episode 开始时从区间 $[0.5, 3.0]$ dB 内随机采样阴影衰落标准差 $\sigma_{\text{dB}}^{\text{env}}$, 而本文方法内部的贝叶斯滤波器则以固定假设值 $\sigma_{\text{dB}} = 2.8$ dB 进行似然计算, 由此产生的模型失配进一步检验了系统的鲁棒性。每次仿真持续 $T = 220$ 个时间步, 辐射源随机执行直线、锯齿形、矩形与随机游走等典型机动轨迹, 双方初始位置随机生成。所有实验在配备 NVIDIA RTX 3090 GPU 的服务器上完成。在全部测试轨迹上统计, 单时间步端到端推理耗时 (涵盖感知层特征提取、认知层双链推断更新与决策层策略评估) 均值为 11.6 ms, 标准差为 2.3 ms。

表 1 仿真参数设置

Table 1 Simulation parameter settings

参数	符号	数值
栅格空间尺寸	$H \times W$	64×64
单步飞行步长	Δ_s	1 m
追踪捕获半径	r_{track}	8 m
环境阴影衰落标准差	$\sigma_{\text{dB}}^{\text{env}}$	$[0.5, 3.0]$ dB
滤波器假设噪声标准差	σ_{dB}	2.8 dB
单次仿真时间步	T	220
参数粒子数	K	12
IMM 子滤波器数	M	3
MPC 前瞻窗口	N_h	6
时间折扣因子	γ	0.9
ESS 复苏阈值	τ_{ess}	0.45
复苏替换比例	—	35%
均匀采样粒子比例	—	25%
IMM 切换概率范围	$[\pi_{\text{min}}, \pi_{\text{max}}]$	$[0.005, 0.08]$
熵缩放因子	η	1.2
越界惩罚系数	λ_{out}	160
转弯惩罚系数	λ_{turn}	0.06
探测前瞻步数	N_{probe}	1
探测折扣因子	γ_{probe}	0.9
探测激活阈值	H_{safe}	2.7
探测激活阈值	H_{safe}	2.7

为评估不同先验条件下的适应能力,每种方案均在两种初始化模式下分别运行。无先验初始化模式中传播参数 θ 完全未知,系统从广域分布冷启动;真值初始化模式中传播参数直接设为仿真器的真实值,代表参数已知的理想上界。两种模式的性能差距直接反映了各方法对参数不确定性的鲁棒程度。为全面评估所提方法的性能优势,本文选取以下基准方案进行对比。

(1) 粒子滤波 SIR: 采用标准序贯重要性重采样粒子滤波器进行目标状态估计,决策层使用贪心策略直接飞向 MAP 估计位置。该方案不具备参数在线估计与主动探测能力,代表经典被动滤波范式的性能基准。

(2) POMCP 类规划: 认知层沿用本文的双链 RBPF-IMM 推断架构,决策层替换为基于蒙特卡罗树搜索的前瞻规划算法,通过多步仿真评估候选动作的长期期望收益,代表经典 POMDP 在线求解方法。

(3) EKF 贪心: 采用扩展卡尔曼滤波器在连续状态空间中维护目标位置的高斯后验估计,决策层使用贪心策略。该方案以单高斯假设近似后验,代表轻量级连续域滤波方法。

(4) 随机策略: 认知层使用本文的完整推断架构,决策层在每个时间步从 A 中等概率随机选取动作。该方案用于量化主动决策的性能增益,同时验证认知层在缺乏合理决策支撑时的退化程度。

图 3 给出了参数先验网络与排序门控网络的训练损失曲线。参数先验推断网络与排序门控网络的训练数据均来自仿真平台生成的随机轨迹集合,训练过程独立于测试场景。部署阶段两个网络的参数保持冻结,系统以纯在线推理模式运行,单次时间步的端到端推理耗时约 12 ms。两个网络均在前 10 轮内完成损失的快速下降,随后平稳收敛。参数先验网络的训练与验证损失全程保持紧密贴合,未出现验证损失反弹,表明其在参数经验分布提取任务上不存在过拟合。排序门控网络的收敛速度略快,约 15 轮后训练间距即趋于稳定,这与其输入维度较低且候选空间仅含三个离散策略分支的结构特点一致。两个网络均展现出良好的泛化性能,为部署阶段的实时闭环推断提供了可靠的模型基础。

图 4 给出了五种方案的 AUC 归一化指标总体对比。本文方法在无先验初始化条件下取得了最高 AUC 值,且显著领先于其余四种基线。本文方法在参数完全未知时的表现已超越粒子滤波 SIR 与 EKF 贪心在真值初始化下的性能,说明感知层的在线参数推断与认知层的双链解耦推断能够有效弥补参数缺失带来的信息损失,决策层的分层规划进一步将认知

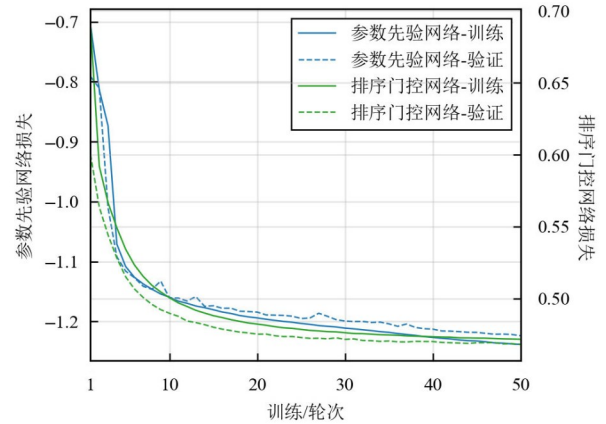


图 3 训练损失曲线

Figure 3 Training loss curves

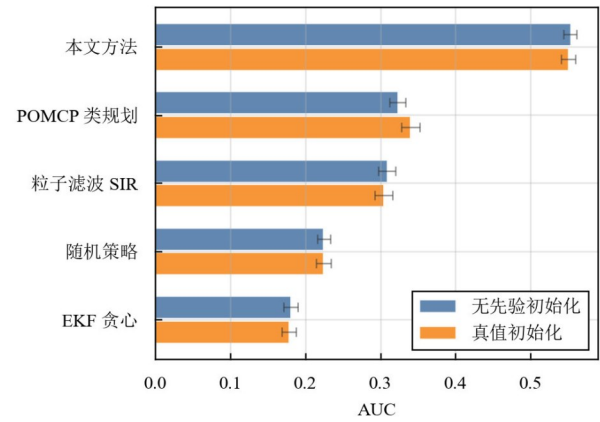


图 4 AUC 归一化指标总体对比

Figure 4 Overall comparison of normalized AUC across methods

优势转化为物理追踪收益。POMCP 类规划虽具备多步前瞻搜索能力,但 AUC 仍明显低于本文方法,反映了蒙特卡罗树搜索在高维信念空间中的采样效率瓶颈。随机策略在两种模式下均表现最差且初始化模式间几乎无差异,表明缺乏合理决策引导时认知层的推断优势无法转化为有效的追踪行为。

图 5 给出了各方案的平均目标距离对比。本文方法在无先验模式下取得了最低的平均距离,且优于粒子滤波 SIR 在真值初始化下的结果,佐证了推断与决策联合优化的有效性。POMCP 类规划位列第二,但与本文方法仍有显著差距。随机策略与 EKF 贪心的平均距离均超过 30 格,后者在两种初始化模式下差异较小,源于单高斯假设无法表征目标机动产生的多峰位置分布,导致贪心策略频繁飞向错误的概率重心。

为揭示追踪性能的动态演化过程,图 6 展示了各方案在无先验模式下平均目标距离随时间步的变化曲线。本文方法在约前 30 步的冷启动阶段距离较高,对应于感知层输出高方差参数先验、决策层偏向

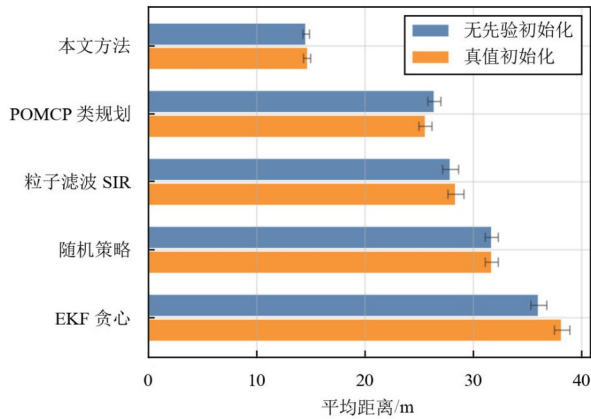


图5 平均目标距离对比

Figure 5 Comparison of average target distance across methods

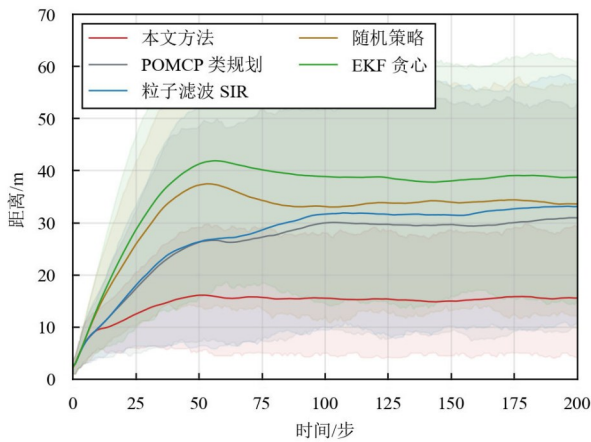
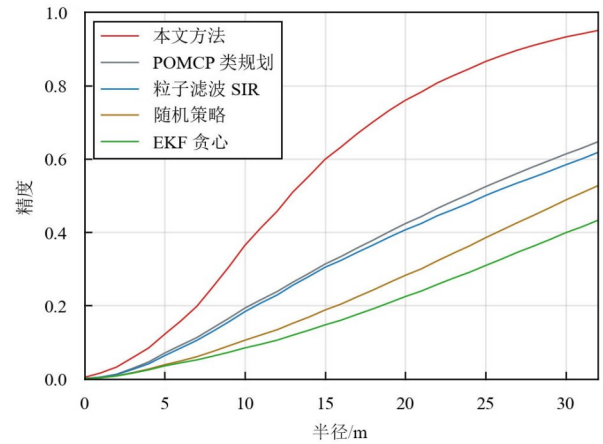


图6 对比平均目标距离随时间演化

Figure 6 Evolution of average target distance over time steps

探测策略的系统状态。随后距离迅速下降并在 50 步左右收敛至稳定的低距离区间,体现了从主动探索到深度利用的阶段转变。其物理机制在于早期的大尺度空间机动为参数粒子提供了充足的观测分集,使双链推断迅速锁定联合后验,信念熵的下降随即触发决策层切换至追踪策略。粒子滤波 SIR 在初期与本文方法差距不大,但收敛后的稳态距离明显更高且持续波动,反映了被动滤波在目标机动时无法通过调整飞行方向恢复追踪的局限。其余方案的距离曲线始终维持在高位,未展现出有效的收敛特征。

为更精细地刻画空间分辨能力,图 7 给出了各方案在无先验模式下的 Precision@ r 曲线。本文方法在所有半径阈值上均位于最上方,且优势在中等半径区间尤为显著: $r=16$ 格时本文方法精度约 0.63,粒子滤波 SIR 与 POMCP 类规划均约 0.33, EKF 贪心仅约 0.16。 $r=8$ 格处虚线对应预设捕获半径 r_{track} ,本文方法在该处精度约 0.25。在小半径区间,本文方法的曲线斜率明显更高,说明在要求精细逼近目标的严苛条

图7 Precision@ r 对比曲线Figure 7 Comparison curves of Precision@ r across methods

件下同样具备显著的定位能力。

图 8 聚焦于严格半径精度 Precision@4.0 格的热力图,将评估标准收紧至捕获半径的一半。本文方法在无先验与真值模式下分别达到 0.085 与 0.086,两者几乎无差异,表明感知层的在线参数推断已能充分替代真实参数先验。本文方法的无先验精度约为粒子滤波 SIR 真值模式的两倍、POMCP 类规划真值模式的 1.67 倍、EKF 贪心真值模式的近三倍。其机制在于主动探测策略系统性地引导无人机飞向高信息增益区域,由此获取的观测序列具备更优的空间分集特性,使得联合后验在参数维度和状态维度上同时实现精细收敛。

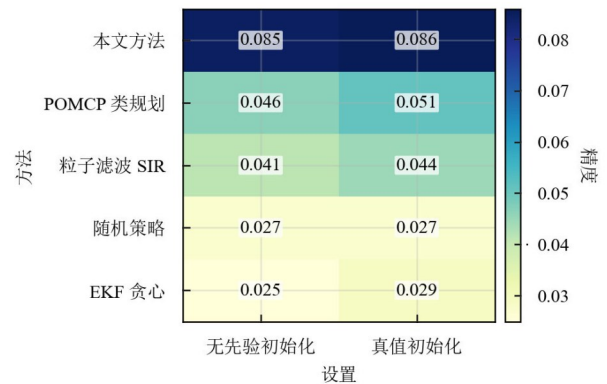


图8 严格半径精度 Precision@4.0 热力图

Figure 8 Heatmap of strict-radius precision (Precision@4.0)

图 9 给出了各方案在捕获半径 r_{track} 内的追踪比例对比。本文方法在无先验模式下达到约 0.35,意味着在 220 步的仿真周期中约有 77 个时间步处于捕获半径以内,考虑到约 30 步的冷启动探索期,稳态阶段的实际追踪效率显著更高。该值同样超越了粒子滤波 SIR 在真值初始化下的表现,验证了主动决策对追踪质量的关键贡献。EKF 贪心与随机策略的追踪比例均低于 0.10,系统大部分时间处于目标丢失状态。

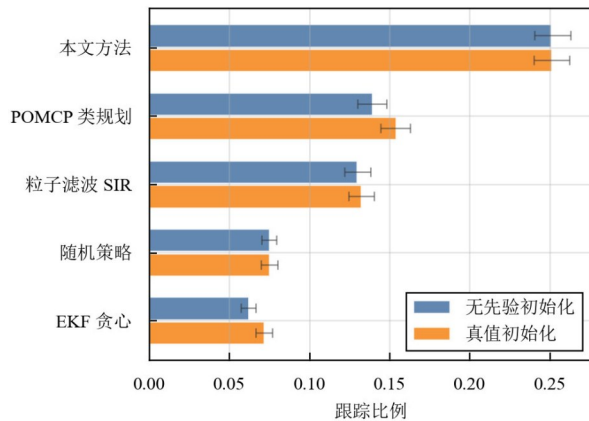


图9 捕获半径内追踪比例

Figure 9 Tracking ratio within the capture radius

图 10 展示了各方案的最长连续捕获占比。相较于追踪比例,该指标对稳定性要求更为严苛,任何一次短暂的目标丢失都会中断连续计数。本文方法在无先验模式下接近 0.18,远高于其余基线,表明系统不仅能快速收敛至目标附近,更能在目标持续机动条件下维持长时间稳定伴飞。该能力源自认知层自适应 IMM 机制与决策层强制探测覆盖规则的协同:当目标突然变向导致信念熵骤升时,IMM 滤波器通过提升模式切换概率快速捕获新运动模式,决策层据此切换至探测策略以恢复信念收敛,从而避免长时间追踪中断。POMCP 类规划位列第二,但连续捕获占比不足本文方法的一半,其余方案均低于 0.05。

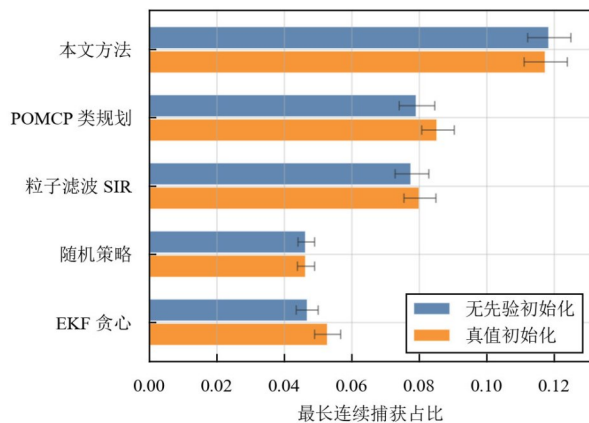


图 10 最长连续捕获占比

Figure 10 Longest continuous capture ratio

图 11 给出了一次典型仿真 episode 的空间轨迹与距离演化曲线。左图中无人机在起始阶段执行了较大范围的发散式探索,这是决策层在高信念熵驱动下偏向探测策略的外在表现。随着观测分集的累积,无人机轨迹逐渐收敛至目标附近并进入紧密伴飞模式,两条轨迹在后半段呈现高度平行的几何关系。右图

的距离时序曲线显示,系统在约 30 步后距离迅速下降至捕获半径以内,此后虽因目标机动偶有短暂跳出,但均在数步内恢复,最终取得了 0.717 的追踪占比与 9.14 m 的平均距离。该结果与 3.4 节所分析的冷启动探索、过渡收敛与稳态追踪三阶段行为高度吻合。

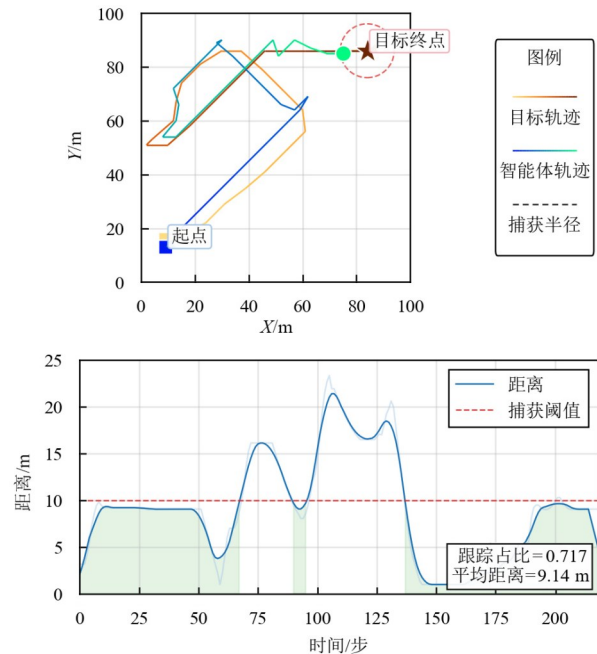


图 11 典型追踪过程可视化

Figure 11 Visualization of a typical tracking episode

4 结论

本文针对传播参数未知的机动辐射源单机追踪难题,构建了以不确定度双向传递为纽带的感知、推断与决策闭环具身追踪系统。该系统以双链解耦推断为核心,通过 Rao-Blackwellized 粒子滤波与交互多模型的嵌套协同,高效实现了参数与状态的联合在线收敛;在此基础上,决策层利用排序门控网络对动作空间进行降维优选,使无人机能依据信念不确定度,自发完成从广域探索到紧密伴飞的策略转换。仿真结果验证了该协同架构在恶劣感知条件下的鲁棒性与性能优势。未来研究将突破当前的一阶马尔可夫与参数准静态等基础前提,重点拓展系统在传播参数快变、目标长时序相关及空间物理约束等复杂场景下的追踪能力。

参考文献

- [1] Zeng Y, Wu Q Q, Zhang R. Accessing from the sky: A tutorial on UAV communications for 5G and beyond[J]. Proceedings of the IEEE, 2019, 107(12): 2327-2375.

- [2] 安建平, 李建国, 于季弘, 等. 空天通信网络关键技术综述[J]. 电子学报, 2022, 50(2): 470-479.
An Jianping, Li Jianguo, Yu Jihong, et al. Key technologies of space-air-ground communication networks: A survey[J]. Acta Electronica Sinica, 2022, 50(2): 470-479. (in Chinese)
- [3] Obeidat H, Shuaieb W, Obeidat O, et al. A review of indoor localization techniques and wireless technologies[J]. Wireless Personal Communications, 2021, 119(1): 289-327.
- [4] Zhou F H, Wang C Y, Wu G Y, et al. Accurate spectrum map construction for spectrum management through intelligent frequency-spatial reasoning[J]. IEEE Transactions on Communications, 2023, 71(7): 3932-3945.
- [5] Coluccia A, Fascista A. On the hybrid TOA/RSS range estimation in wireless sensor networks[J]. IEEE Transactions on Wireless Communications, 2018, 17(1): 361-371.
- [6] Wojcicki P, Zientarski T, Charytanowicz M, et al. Estimation of the path-loss exponent by Bayesian filtering method[J]. Sensors, 2021, 21(6): 1934.
- [7] Vaghefi R M, Gholami M R, Buehrer R M, et al. Cooperative received signal strength-based sensor localization with unknown transmit powers[J]. IEEE Transactions on Signal Processing, 2013, 61(6): 1389-1403.
- [8] Zholamanov B, Saymbetov A, Nurgaliyev M, et al. RSSI fingerprint-based indoor localization solutions using machine learning algorithms: A comprehensive review[J]. Smart Cities, 2025, 8(5): 153.
- [9] Liu J Y, Liu X D, Liang H T, et al. A data-and-semantic dual-driven intelligent inference framework for simultaneously spectrum map construction and signal source localization[J]. IEEE Internet of Things Journal, 2025, 12(12): 18750-18764.
- [10] 段世红, 何昊, 徐诚, 等. 未知环境下基于深度序列蒙特卡罗树搜索的信源导航方法[J]. 电子学报, 2022, 50(7): 1744-1752.
Duan Shihong, He Hao, Xu Cheng, et al. DS-MCTS: A deep sequential Monte-Carlo tree search method for source navigation in unknown environments[J]. Acta Electronica Sinica, 2022, 50(7): 1744-1752. (in Chinese)
- [11] Zhou F H, Liu C Y, Zhang H, et al. SpectrumFM: A foundation model for intelligent spectrum management[J]. IEEE Journal on Selected Areas in Communications, 2026, 44: 4471-4488.
- [12] Cheng C, Tournet J Y, Lu X D. A Rao-blackwellized particle filter with variational inference for state estimation with measurement model uncertainties[J]. IEEE Access, 2020, 8: 55665-55675.
- [13] Li X H, Lu B, Li Y X, et al. Adaptive interacting multiple model for underwater maneuvering target tracking with one-step randomly delayed measurements[J]. Ocean Engineering, 2023, 280: 114933.
- [14] Popović M, Vidal-Calleja T, Hitz G, et al. An informative path planning framework for UAV-based terrain monitoring[J]. Autonomous Robots, 2020, 44(6): 889-911.
- [15] Charrow B, Liu S K, Kumar V, et al. Information-theoretic mapping using cauchy-schwarz quadratic mutual information[C]//2015 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2015: 4791-4798.
- [16] Hoffmann F, Charlish A, Ritchie M, et al. Policy rollout action selection in continuous domains for sensor path planning[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(4): 2247-2264.
- [17] Carvalho Chanel C, Teichteil-Königsbuch F, Lesire C. Multi-target detection and recognition by UAVs using online POMDPs[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2013, 27(1): 1381-1387.
- [18] Goel H, Lipschitz L J, Agarwal S, et al. Reinforcement learning for agile active target sensing with a UAV[PP/OL]. V1. arXiv (2022-12-16) [2026-05-18]. <https://doi.org/10.48550/arXiv.2212.08214>.
- [19] Xia Z Y, Du J, Wang J J, et al. Multi-agent reinforcement learning aided intelligent UAV swarm for target tracking[J]. IEEE Transactions on Vehicular Technology, 2022, 71(1): 931-945.
- [20] Sunberg Z, Kochenderfer M. Online algorithms for POMDPs with continuous state, action, and observation spaces[J]. Proceedings of the International Conference on Automated Planning and Scheduling, 2018, 28: 259-263.
- [21] Hu J Z, Zhang H L, Song L Y, et al. Reinforcement learning for a cellular Internet of UAVs: Protocol design, trajectory control, and resource management[J]. IEEE Wireless Communications, 2020, 27(1): 116-123.
- [22] Bajcsy R, Aloimonos Y, Tsotsos J K. Revisiting active perception[J]. Autonomous Robots, 2018, 42(2): 177-196.
- [23] Liang H T, Diao Y H, Wu Y H, et al. Synergetic empowerment: Wireless communications meets embodied intelligence[PP/OL]. V2. arXiv (2026-05-26) [2026-05-18]. <https://doi.org/10.48550/arXiv.2509.10481>.
- [24] Liu H P, Guo D, Cangelosi A. Embodied intelligence: A synergy of morphology, action, perception and learning

[J]. ACM Computing Surveys, 2025, 57(7): 1-36.

- [25] Duan J F, Yu S, Tan H L, et al. A survey of embodied AI: From simulators to research tasks[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022,

6(2): 230-244.

- [26] An Z L, Yu X Q, Wang C, et al. Embodied intelligence: Recent advances and future perspectives[J]. The Innovation Informatics, 2025, 1(1): 100008.

作者简介



梁宏韬 男,2000年10月出生于甘肃省白银市。现为南京航空航天大学电子信息工程学院博士研究生。主要研究方向为低空具身智能、多模态融合与电磁频谱具身智能。

E-mail: ceie.lht@nuaa.edu.cn



刘晓东 男,1993年12月出生于江西省赣州市。现为南昌大学信息工程学院讲师。主要研究方向为大模型赋能电磁频谱管理,可见光通信感知照明一体化系统。

E-mail: xiaodongliu@ncu.edu.cn



付杨美子 女,2004年11月出生于江西省南昌市。现为南昌大学信息工程学院本科生,主要研究方向为频谱信号分析。

E-mail: 6105122008@email.ncu.edu.cn



吴启晖 男,1970年12月出生于安徽省黄山市。现为南京航空航天大学特聘教授、副校长,教育部“长江学者”特聘教授,南京航空航天大学电子信息工程学院教授、博导。主要研究方向为认知科学与应用、电磁空间频谱认知智能管控、无人机认知集群。

E-mail: wuqihui@nuaa.edu.cn



万奕尧 男,1997年9月出生于湖北省宜昌市。现为南京航空航天大学博士研究生。研究方向:深度学习、多模态融合。

E-mail: yiyawan@nuaa.edu.cn