

文章编号:1671-1637(2026)04-0121-13

# 面向林区公路高风险路段无人机自适应 巡航轨迹规划方法

陈德启,张淑慧,张文会\*,蒋贤才

(东北林业大学 土木与交通学院,黑龙江 哈尔滨 150040)

**摘要:**为降低因森林遮蔽所引发的林区道路交通事故风险,提出了一种融合状态不确定性的深度强化学习无人机自适应巡航模型。在感知层面,设计了一种适用于林区公路场景的自适应无迹卡尔曼滤波(UKF)算法,来应对全球定位系统信号缺失;在决策层面,构建了基于状态不确定性感知的柔性演员-评论家(SUA-SAC)算法,将UKF输出的状态估计值及其协方差作为网络输入,使SUA-SAC算法能够学习到对状态估计具有更优鲁棒性的控制策略。结果表明:在训练效率方面,SUA-SAC算法的收敛速度相较于基线SAC算法和近端策略优化算法分别提升了约50%和60%;在多场景测试中,相较于SAC算法,SUA-SAC算法在无干扰、动态遮挡和强风干扰场景下的平均跟踪误差分别降低了67%、61%和66%;在定位信号缺失达20s的测试中,SUA-SAC算法的跟踪误差所受影响较小。SUA-SAC算法能够提高无人机在林区复杂路况下的轨迹跟踪精度、飞行稳定性以及任务成功率,有助于提升林区公路的交通安全水平。

**关键词:**低空交通;交通安全;林区公路;无人机巡航;SUA-SAC算法

**中图分类号:**U8 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2026.168

## Method for UAV adaptive cruising trajectory planning for high-risk road sections of forest roads

CHEN De-qi, ZHANG Shu-hui, ZHANG Wen-hui\*, JIANG Xian-cai

(School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, Heilongjiang, China)

**Abstract:** To mitigate forest road traffic accident risks caused by canopy obstruction, a deep reinforcement learning (DRL)-based adaptive cruising model for UAVs that accounts for state uncertainty was proposed. At the perception stage, an adaptive unscented Kalman filter (UKF) tailored to forest road scenarios was designed to address global positioning system (GPS) signal loss. At the decision-making stage, a state uncertainty-aware soft actor-critic (SUA-SAC) algorithm was developed, where UKF-derived state estimates and their covariance were used as network inputs, enabling SUA-SAC to learn control strategies that are more robust to state estimation. The results show that, in terms of training efficiency, the convergence speed of SUA-

**出版历程:**2025-10-10 收稿,2025-12-11 修回,2026-01-23 录用

**基金项目:**黑龙江省哲学社会科学规划项目(23GLC022);国家自然科学基金项目(52572369)

**作者简介:**陈德启(1990-),男,黑龙江哈尔滨人,讲师,工学博士,E-mail:chendeqi@nefu.edu.cn.

\***通信作者:**张文会(1978-),男,黑龙江哈尔滨人,教授,博士生导师,工学博士,E-mail:zhangwenhui@nefu.edu.cn.

**引用格式:**陈德启,张淑慧,张文会,等.面向林区公路高风险路段无人机自适应巡航轨迹规划方法[J].交通运输工程学报,2026,26(4):121-133.

**Citation:** CHEN De-qi, ZHANG Shu-hui, ZHANG Wen-hui, et al. Method for UAV adaptive cruising trajectory planning for high-risk road sections of forest roads[J]. Journal of Traffic and Transportation Engineering, 2026, 26(4): 121-133.

SAC is improved by approximately 50% and 60% compared with the baseline SAC algorithm and proximal policy optimization algorithm. In multi-scenario tests, compared with the SAC algorithm, SUA-SAC reduces the average tracking error by 67%, 61%, and 66% in scenarios without interference, dynamic occlusion, and strong wind interference, respectively. In tests involving positioning signal loss lasting up to 20 s, the tracking error of SUA-SAC is less affected. Overall, SUA-SAC improves UAV trajectory tracking accuracy, flight stability, and mission success rate under complex forest road conditions, contributing to the enhancement of traffic safety on forest roads.

**Keywords:** low-altitude traffic; traffic safety; forest road; UAV cruising; SUA-SAC algorithm

**Publication history:** Received 2025-10-10; Received in revised form 2025-12-11; Accepted 2026-01-23

**Funding:** Annual Philosophy and Social Science Foundation of Heilongjiang Province (23GLC022); National Natural Science Foundation of China (52572369)

\* **Corresponding author:** ZHANG Wen-hui, professor, PhD, E-mail: zhangwenhui@nefu.edu.cn.

## 0 引 言

林区公路受茂密树冠视线遮蔽、复杂的道路形态、以及林间多变的强风影响,形成了复杂且充满不确定性的交通环境,存在较多的交通事故高风险路段。林区道路交通安全巡查若单纯依赖传统人工巡检进行路况监测与事故处理,不仅效率低下而且成本高昂<sup>[1]</sup>。无人机凭借其独特的机动能力、灵活的部署方式和广阔的作业视角<sup>[2]</sup>,为林区公路的智能化运维提供了新的解决方案。

无人机在林区自主飞行是一个复杂的序贯决策问题,不仅要规划出满足任务要求的最优轨迹,更要在林区非结构化的公路环境下对轨迹进行精确跟踪<sup>[3]</sup>。这一过程的成败,直接取决于 2 个难题:一是在定位信号(Global Positioning System, GPS)丢失时,获得连续、可靠的状态信息;二是在状态信息缺失的情况下,做出安全、精准的飞行控制决策。

为解决无人机高精度轨迹跟踪问题,大量研究工作首先从经典与现代控制理论中寻求方案。比例-积分-微分控制及变体因其结构简单、易于实现,在无人机飞行控制中得到了基础性应用<sup>[4-5]</sup>。为应对无人机飞行的非线性特性,研究人员进一步引入了非线性控制方法。如 Zhao 等<sup>[6]</sup>采用自适应快速终端滑模控制器,有效抑制了外部扰动。模型预测控制则通过在线优化,在处理约束和优化性能方面展现出强大能力<sup>[7-8]</sup>。然而,上述模型普遍存在一定的局限性。其通常假定能够获取精确的系统状态反馈,但在森林环境下,尤其当定位信号丢失时,这些控制器可能因错误的反馈信息而失稳。

深度强化学习(Deep Reinforcement Learning,

DRL)是一种“感知-决策”一体化的端到端学习方法<sup>[9]</sup>,能够通过与环境的直接交互和试错<sup>[10]</sup>,自主学习从高维传感器输入到最优控制输出的复杂映射。早期的 DRL 算法,如深度 Q 网络,主要处理离散动作空间,而无人机控制是典型的连续域问题<sup>[11]</sup>。因此,无人机轨迹控制研究转向了能够处理连续动作空间的演员-评论家架构、深度确定性策略梯度<sup>[12-13]</sup>、近端策略优化(Proximal Policy Optimization, PPO)等算法<sup>[14-15]</sup>。Ma 等<sup>[16]</sup>利用 DRL 成功训练了无人机在风扰环境下的跟踪控制。其中,柔性演员-评论家(Soft Actor-critic, SAC)算法因其卓越的样本效率和在探索-利用平衡上的出色表现,被公认为是解决连续控制问题的最先进方法之一<sup>[17-18]</sup>。

然而,现有研究在应用 DRL 算法时,大多隐含了一个重要假设:智能体能够获取相对确定的环境状态,但在林区公路场景下,GPS 信号缺失导致的状态估计不确定性,成为了制约标准 DRL 算法性能的瓶颈<sup>[19-20]</sup>。当状态估计器输出一个置信度较低的状态时,标准 SAC 算法无法区分优劣,仍会将其作为一个确定的输入进行决策。近年来,一些前沿研究开始关注决策过程中的不确定性问题<sup>[21]</sup>。部分工作通过在模型中引入贝叶斯推断或高斯过程<sup>[22]</sup>,来量化模型的不确定性<sup>[23-24]</sup>,另一些工作则尝试让智能体直接从带噪声的观测中学习<sup>[25-26]</sup>。此外,在环境风险规避与多机协同规划层面,相关工作通过融合人员伤亡、噪声干扰及风场影响等多源风险信息,构建了动态风险地图模型,并配合路径重规划与起点等待的组合策略实现了复杂环境下多机冲突的有效消解<sup>[27-28]</sup>。这类方法通过先验规划有效规避了

外部环境风险,但在林区公路等高遮蔽场景下,由 GPS 信号缺失引发的智能体自身感知层面的“状态估计不确定性”往往更具随机性与瞬态性。因此,目前仍缺少一种能将时变状态估计不确定性与 DRL 决策过程进行深度显式耦合的无人机飞行控制框架。

## 1 问题描述与模型建立

本节从 3 个层面展开:一是建立无人机的运动学模型,以精确描述其在三维空间中的动态行为;二是设计基于无迹卡尔曼滤波的自适应状态估计器,以从带噪声的传感器数据中获取精确、鲁棒的无人机状态;三是将轨迹跟踪任务构建为马尔可夫决策过程(Markov Decision Process, MDP)框架,从而将复杂的连续控制问题转化为一个标准的、可由强化学习算法求解的序贯决策问题。

### 1.1 无人机运动学建模

由于真实的四旋翼无人机是一个复杂的非线性、强耦合系统,为了在仿真环境中进行控制方法的研究,现作出如下假设。

(1)将无人机机身视为 1 个质量恒定、外形尺寸不发生改变的刚体。在飞行过程中,不考虑因高速旋转或碰撞导致的结构形变。

(2)无人机结构对称,其几何中心与质心重合。

(3)忽略地球自转及复杂的空气动力学效应干扰。

(4)假设无人机的 4 个电机响应速度极快,能够瞬时输出期望的力,忽略电机和螺旋桨的动态响应过程。

首先,定义无人机在三维空间中的运动状态向量  $\mathbf{X}$  为

$$\mathbf{X} = (\mathbf{x}_E, \boldsymbol{\eta}, \mathbf{V}_B, \boldsymbol{\omega}_B)^T \quad (1)$$

式中: $\mathbf{x}_E$  为无人机质心在地面坐标系下的三维坐标, $\mathbf{x}_E = (x, y, z)^T$ ,其中, $x, y, z$  分别为轴上的位置坐标; $\boldsymbol{\eta}$  为描述无人机自身姿态的 3 个角度, $\boldsymbol{\eta} = (\phi, \xi, \Psi)^T$ ,其中, $\phi, \xi, \Psi$  分别为滚转角、俯仰角、偏航角; $\mathbf{V}_B$  为无人机在自身坐标系下的速度, $\mathbf{V}_B = (u, v, w)^T$ ,其中, $u, v, w$  分别为无人机在机体坐标系下,沿其前、右、下 3 个轴向的线速度分量; $\boldsymbol{\omega}_B = (\omega_x, \omega_y, \omega_z)^T$  为机体坐标系下的角速度向量。

其次,定义系统的控制输入向量  $\mathbf{U}$  为

$$\mathbf{U} = (T, \tau_\phi, \tau_\xi, \tau_\Psi)^T \quad (2)$$

式中: $T$  为 4 个旋翼产生的总推力,使无人机克服重力并进行升降; $\tau_\phi, \tau_\xi, \tau_\Psi$  分别为无人机旋翼间转速差产生的滚转、俯仰、偏航控制力矩。

最后,建立状态方程  $\dot{\mathbf{X}} = f(\mathbf{X}, \mathbf{U})$ 。 $\dot{\mathbf{X}}$  为  $\mathbf{X}$  的变化率,其为

$$\dot{\mathbf{X}} = (\dot{\mathbf{x}}_E, \dot{\boldsymbol{\eta}}, \dot{\mathbf{V}}_B, \dot{\boldsymbol{\omega}}_B)^T \quad (3)$$

$$\dot{\mathbf{x}}_E = \mathbf{C}_B^E(\boldsymbol{\eta})\mathbf{V}_B \quad (4)$$

$$\dot{\boldsymbol{\eta}} = \mathbf{W}(\boldsymbol{\eta})\boldsymbol{\omega}_B \quad (5)$$

$$\dot{\mathbf{V}}_B = -\boldsymbol{\Omega}(\boldsymbol{\omega}_B)\mathbf{V}_B + \frac{\mathbf{F}_e}{m} \quad (6)$$

$$\dot{\boldsymbol{\omega}}_B = \mathbf{J}^{-1}[\boldsymbol{\tau}_e - \boldsymbol{\Omega}(\boldsymbol{\omega}_B)\mathbf{J}\boldsymbol{\omega}_B] \quad (7)$$

式中: $\dot{\mathbf{x}}_E$  为位置变化率; $\dot{\boldsymbol{\eta}}$  为姿态变化率; $\dot{\mathbf{V}}_B$  为线速度变化率; $\dot{\boldsymbol{\omega}}_B$  为角速度变化率; $\mathbf{C}_B^E(\boldsymbol{\eta}) \in \mathbb{R}^{3 \times 3}$  为从机体坐标系到惯性坐标系的旋转矩阵; $\mathbf{W}(\boldsymbol{\eta}) \in \mathbb{R}^{3 \times 3}$  为将机体角速度转换为欧拉角变化率的转换矩阵; $\boldsymbol{\Omega}(\boldsymbol{\omega}_B)$  为由角速度向量  $\boldsymbol{\omega}_B$  构造的反对称矩阵算子; $m$  为无人机的总质量; $\mathbf{F}_e$  为无人机上的外部合力向量; $\mathbf{J}$  为转动惯量矩阵; $\boldsymbol{\tau}_e$  为作用在无人机上的外部合力矩向量。

### 1.2 基于无迹卡尔曼滤波的状态估计

设计了鲁棒的状态估计器,从带噪声的测量中提取无人机真实状态的最优估计,采用无迹卡尔曼滤波(Unscented Kalman Filter, UKF)作为核心估计算法,通过无迹变换来处理非线性,避免了线性化误差和雅可比矩阵的复杂计算,能够以更高阶的精度逼近非线性系统的概率分布,从而获得更准确、更鲁棒的状态估计结果。

#### 1.2.1 系统模型

首先将建立的连续时间动力学模型转换为离散时间形式,如下所示

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{q}_k \quad (8)$$

式中: $\mathbf{x}_k$  为  $k$  时刻的系统状态向量; $\mathbf{u}_k$  为  $k$  时刻的控制输入向量; $f(\cdot)$  为非线性状态转移函数; $\mathbf{q}_k$  为  $k$  时刻的过程噪声向量。

传感器的测量过程可建模为

$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{n}_k \quad (9)$$

式中: $\mathbf{y}_k$  为  $k$  时刻的传感器测量向量; $h(\cdot)$  为非线性测量函数,将系统状态向量  $\mathbf{x}_k$  映射到测量空间; $\mathbf{n}_k$  为  $k$  时刻测量噪声向量。

#### 1.2.2 林区公路场景下的 GPS 噪声模型

在林区公路的特殊环境下,无人机在树冠下方穿行时,GPS 信号会因遮挡和多径效应而急剧恶化,导致其测量噪声显著增大。传统的 UKF 使用固定的测量噪声协方差矩阵  $\mathbf{G}$ ,无法适应这种随时变化的环境特性。为了解决这一问题,本文提出自适应测量噪声协方差矩阵(Adaptive Measurement Noise Covariance, AMNC)的策略,根据 GPS 信号

的可靠性,动态调整协方差矩阵  $\mathbf{G}$  中对应 GPS 测量项的参数。

首先定义 1 个 GPS 信号可靠性因子  $\lambda_k \in [0, 1]$ , 用于量化当前时刻 GPS 定位的置信度。考虑到水平精度因子 (Horizontal Dilution of Precision, HDOP) 能够综合反映可见卫星数量与几何分布对定位精度的影响,选取 HDOP 作为计算基准,并构建基于 HDOP 的归一化映射机制来实时解算  $\lambda_k$ 。具体实现方式为:设定 HDOP 的理想基准值对应最高置信度 ( $\lambda_k = 1$ ),并设定信号失效的截止阈值。当实时 HDOP 值处于两者之间时, $\lambda_k$  随 HDOP 值的增加呈非线性单调递减。该机制确保了滤波器能够根据接收机输出的信号质量指标,平滑地调整对测量数据的信任权重。基于此, $k$  时刻的自适应测量噪声协方差矩阵  $\mathbf{G}_k$  可建模为

$$\mathbf{G}_k = \frac{\mathbf{G}_{\text{base}}}{\lambda_k + \epsilon} = \frac{1}{\lambda_k + \epsilon} \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_z^2) \quad (10)$$

式中: $\text{diag}(\sigma_x^2, \sigma_y^2, \sigma_z^2)$  为理想条件下的基础测量噪声协方差矩阵, $\sigma_x^2, \sigma_y^2, \sigma_z^2$  分别对应  $x, y, z$  轴的位置测量方差; $\epsilon$  为一个极小的正常数,即  $\epsilon = 1.0 \times 10^{-6}$ , 添加在分母中以确保当可靠性因子  $\lambda_k$  接近 0 时  $\mathbf{G}_k$  依然存在意义,此时  $\mathbf{G}_k$  趋于无穷大,从而使滤波器在更新阶段自动降低对 GPS 观测值的权重。

通过建立系统过程模型、测量模型以及针对性的自适应噪声模型,UKF 算法在每个离散时间步  $k$  都会执行 1 次预测与更新循环。滤波器以内插得到的  $\mathbf{G}_k$  作为观测噪声的统计特性,最终输出无人机在  $k$  时刻的后验状态估计向量  $\mathbf{x}_k$ ,以及反映该估计不确定性的后验状态协方差矩阵  $\xi_k$ 。这 3 个指标为后续的强化学习决策模块提供了关键的输入信息。

### 1.3 马尔可夫决策过程

MDP 通常由五元组  $\{S, A, P, R, \gamma\}$  定义,其中  $S$  为状态空间, $A$  为动作空间, $P$  为状态转移概率, $R$  为奖励函数, $\gamma$  为折扣因子。

#### 1.3.1 状态空间

状态空间  $S$  的设计旨在为智能体提供信息,一个设计良好的状态空间应包含无人机自身的物理状态以及其与任务目标相关的状态,因此,在  $k$  时刻,无人机的状态  $\mathbf{s}_k \in S$  定义如下

$$\mathbf{s}_k = (\mathbf{x}_k, \mathbf{e}_k)^\top \quad (11)$$

$$\mathbf{e}_k = \mathbf{x}_{r,k} - \mathbf{x}_k \quad (12)$$

式中: $\mathbf{x}_k$  为无人机后验状态估计向量,包含了无人机的位置、姿态、线速度和角速度信息; $\mathbf{e}_k$  为任务误差向量; $\mathbf{x}_{r,k}$  为参考轨迹点状态向量,定义了在了  $k$  时

刻无人机期望达到的理想状态。

#### 1.3.2 动作空间

动作空间  $A$  定义了智能体可以执行的所有可能操作的集合,在  $k$  时刻,无人机输出的动作  $\mathbf{a}_k \in A$  定义为

$$\mathbf{a}_k = (T_k, \tau_{\phi,k}, \tau_{\xi,k}, \tau_{\psi,k})^\top \quad (13)$$

式中: $T_k$  为施加于无人机的总推力; $\tau_{\phi,k}, \tau_{\xi,k}, \tau_{\psi,k}$  分别为施加于无人机的滚转、俯仰和偏航控制力矩。

设定最大推重比  $\tau = 6^{[27]}$ ,由此确定物理控制量的约束范围如下:总推力  $T_k$  为  $[0, 28]$  N,滚转、俯仰和偏航控制力矩范围均为  $[-0.2, 0.2]$  N·m,以限制角加速度并防止飞行姿态失稳。

#### 1.3.3 奖励函数

奖励函数是引导智能体学习期望行为的关键。为了平衡任务完成效率与飞行安全性, $k$  时刻的奖励函数  $r_k$  表示如下

$$r_k = \omega_p R_p + \omega_v R_v + \omega_c R_c \quad (14)$$

式中: $\omega_p, \omega_v, \omega_c$  分别为位置跟踪奖励  $R_p$ 、速度跟踪奖励  $R_v$ 、控制能耗惩罚  $R_c$  奖励子项的权重系数。

通过网格搜索方法对权重组合进行了微调。首先,考虑到位置跟踪是避撞与沿路飞行的基础,赋予位置权重  $\omega_p$  最高优先级;其次是速度跟踪权重  $\omega_v$ ,以保证飞行的平滑性;最后是能耗惩罚权重  $\omega_c$ ,用于抑制不必要的控制振荡。经过多组预试验对比,最终确定权重参数为: $\omega_p = 1.0, \omega_v = 0.1, \omega_c = 0.01$ 。该组合能够引导智能体在训练初期快速收敛至安全区域,并在后期优化控制平滑度。

位置跟踪奖励  $R_p$  用于惩罚无人机与期望位置之间的偏差,是核心的跟踪性能指标,表示为

$$R_p = -\|\hat{\mathbf{x}}_{p,k} - \mathbf{x}_{p,r,k}\|^2 \quad (15)$$

式中: $\hat{\mathbf{x}}_{p,k}$  为 UKF 估计的无人机位置向量; $\mathbf{x}_{p,r,k}$  为参考轨迹在  $k$  时刻的期望位置向量。

速度跟踪奖励  $R_v$  用于鼓励无人机保持与参考轨迹相近的速度,使跟踪更平滑,表示为

$$R_v = -\|\hat{\mathbf{v}}_k - \mathbf{v}_{r,k}\|^2 \quad (16)$$

式中: $\hat{\mathbf{v}}_k$  为 UKF 估计的无人机机体线速度向量; $\mathbf{v}_{r,k}$  为参考轨迹的线速度向量。

控制能耗惩罚  $R_c$  为

$$R_c = -\|\mathbf{a}_k\|^2 \quad (17)$$

## 2 基于 SUA-SAC 算法的 UAV 控制方法

为应对无人机轨迹跟踪这一高维连续控制任务,提出一种基于状态不确定性感知的 SAC(State Uncertainty-aware SAC, SUA-SAC)算法。其核心

思想是让智能体能够主动感知并利用状态估计的不确定性,从而学习到在不同置信度下采取不同风险偏好的鲁棒控制策略。

### 2.1 算法整体框架

SUA-SAC 算法整体框架如图 1 所示。该框架构成了一个完整的“感知-估计-决策-控制”闭环系

统,其核心是将 UKF 作为状态估计模块,与 SUA-SAC 算法决策模块进行深度耦合。

在每一个控制周期内,智能体的 Actor 网络首先根据当前的不确定性增强状态向量  $s'_k$  输出控制动作向量  $a_k$ ,该动作被施加于训练环境中的无人机模型。

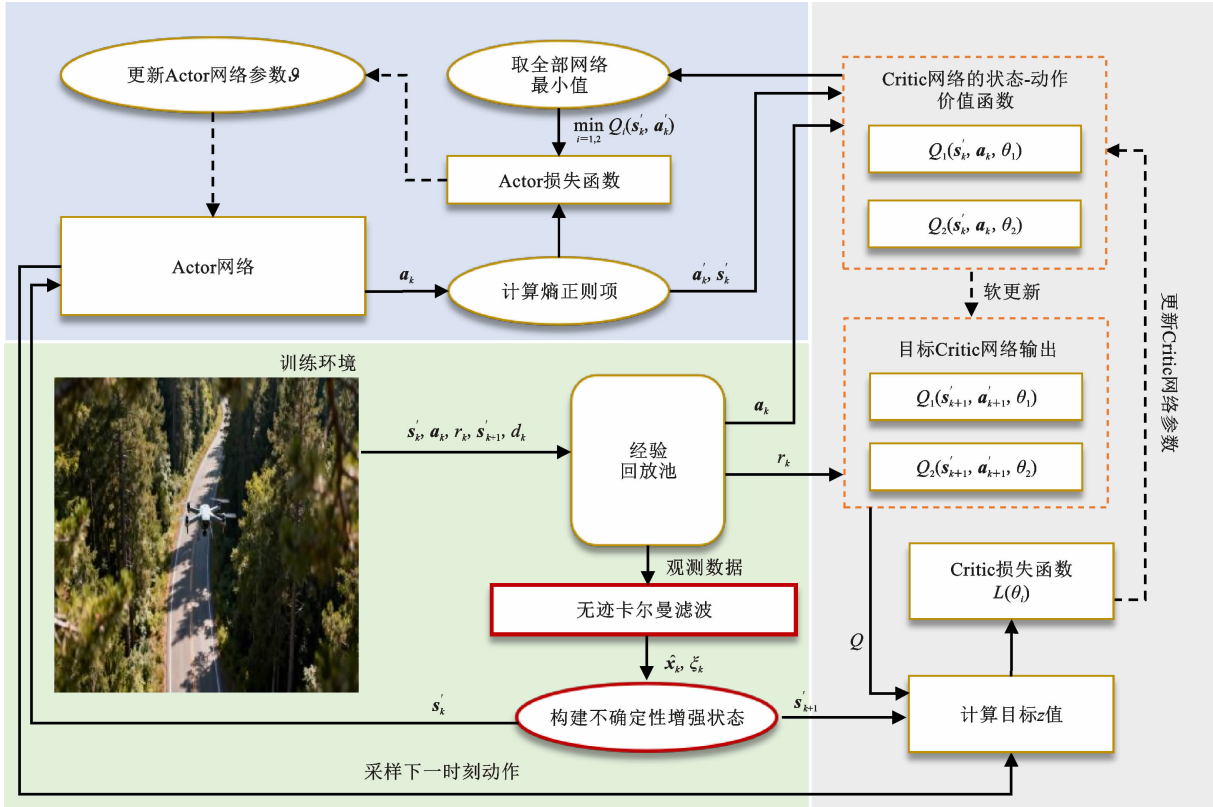


图 1 SUA-SAC 算法框架

Fig.1 SUA-SAC algorithm framework

环境执行该动作后,返回 1 个包含噪声的原始观测量及相应的奖励  $r_k$  和结束标志  $d_k$ 。这些原始观测量并不会直接传递给智能体,而是首先被送入无迹卡尔曼滤波器中。UKF 对带噪声的观测数据进行处理,输出对无人机下一时刻状态的后验状态估计向量  $\hat{x}_{k+1}$  以及反映该估计不确定性的后验状态协方差矩阵  $\xi_{k+1}$ 。随后,这 2 个输出被用于构建新的不确定性增强状态  $s'_{k+1}$ ,并与前一时刻的增强状态、动作、奖励及结束标志共同组成 1 个完整的经验元组  $\{s'_k, a_k, r_k, s'_{k+1}, d_k\}$ ,存入经验回放池中。

在算法的训练阶段,系统会从经验回放池中随机采样一批历史数据。对于 Critic 部分,目标 Q 网络利用采样的下一时刻状态  $s'_{k+1}$  和对应的策略动作  $a_{k+1}$  计算出目标 Q 值,并与主 Q 网络计算的当前 Q 值进行比较,以构造时序差分误差来更新 2 组主 Q 网络参数  $\theta_1$  和  $\theta_2$ ,并通过软更新的方式将主 Q 网络

的参数传递给目标 Q 网络。对于 Actor 部分,策略网络利用采样的当前状态  $s'_k$  产生动作,并通过最小化 Critic 网络输出的 Q 值与策略熵的加权和,来更新其自身参数  $\vartheta$ 。通过 Actor 和 Critic 网络的交替迭代更新,智能体最终能够学习到一个能够根据状态估计值及其不确定性,共同进行最优决策的鲁棒控制策略。

### 2.2 基于 SUA-SAC 算法的智能体控制方法

首先从 UKF 输出的后验协方差矩阵  $\xi_k$  中,提取出能够反映各状态量估计方差的对角线元素,与原始状态向量  $s_k$  进行组合拼接,构成一个不确定性增强的状态向量  $s'_k$ ,作为 SUA-SAC 算法中 Actor 和 Critic 网络的统一输入,其定义如下

$$s'_k = (s_k, \text{diag}(\xi_k))^T \quad (18)$$

式中:  $\text{diag}(\xi_k)$  为提取  $\xi_k$  对角线元素组成向量的函数。

从理论层面分析,协方差向量在决策回路中起到了“自适应风险调节器”的作用。具体而言,协方差矩阵的对角线元素 $\sigma_x^2$ 、 $\sigma_y^2$ 、 $\sigma_z^2$ 量化了当前状态估计的信噪比。在训练过程中,Critic网络通过与环境交互,能够学习到“在高不确定性状态下执行大幅度机动动作会导致回报方差增大或期望回报降低”的价值规律。反向传播后,这一价值判断引导Actor网络形成一种隐式的增益调度机制:当输入的协方差值较小时(高置信度),策略网络倾向于输出高增益动作以实现快速跟踪;当协方差值显著增大时(低置信度),策略网络会自动抑制控制输出的幅值与频率,切换至更保守的稳态保持策略,从而在感知受限条件下最大化长期的鲁棒性与安全性。在SUA-SAC算法中,Actor和Critic网络的更新目标与标准SAC算法保持一致,但所有计算都基于不确定性增强的状态 $s'_k$ 进行,因此,需要对网络进行更新。

### 2.2.1 Critic网络更新

Critic部分由2组Q网络构成以抑制Q值高估。其损失函数 $L(\theta_i)$ 通过最小化软贝尔曼误差进行更新

$$L(\theta_i) = E_{(s'_k, a_k, r_k, s'_{k+1})} \{ [Q_{\theta_i}(s'_k, a_k) - z_k]^2 \} \quad (19)$$

$$z_k = r_k + \gamma \{ \min_{i=1,2} Q_i(s'_{k+1}, a'_{k+1}) - \alpha \ln[\pi_{\vartheta}(a'_{k+1} | s'_{k+1})] \} \quad (20)$$

式中: $L(\theta_i)$ 为第*i*个Q网络的损失函数, $\theta_i$ 为第*i*个Q网络的可训练参数; $Q_{\theta_i}(s'_k, a_k)$ 为第*i*个目标Q网络在增强状态 $s'_k$ 下执行动作 $a_k$ 的价值; $E_{(s'_k, a_k, r_k, s'_{k+1})}$ 为对经验回放池中采样到的增强状态 $s'_k$ 、动作 $a_k$ 、奖励 $r_k$ 与下一时刻增强状态 $s'_{k+1}$ 的期望; $r_k$ 为在*k*时刻获得的即时奖励;折扣因子 $\gamma$ 用于平衡短期奖励与长期奖励; $Q_i(\cdot)$ 为第*i*个目标Q网络; $a'_{k+1}$ 为当前策略 $\pi_{\vartheta}$ 在状态 $s'_{k+1}$ 下采样的动作; $\alpha$ 为熵温度系数。

### 2.2.2 Actor网络更新

Actor网络的目标是学习一个最优策略 $\pi_{\vartheta}$ ,使其在最大化Q值的期望的同时最大化策略熵。其损失函数 $L(\vartheta)$ 为

$$L(\vartheta) = E_{s'_k, a_k \sim \pi_{\vartheta}} \{ \alpha \ln[\pi_{\vartheta}(a_k | s'_k)] - \min_{i=1,2} Q_i(s'_k, a_k) \} \quad (21)$$

式中: $\pi_{\vartheta}(a_k | s'_k)$ 为Actor网络根据增强状态 $s'_k$ 输出动作的概率分布。

## 3 试验与仿真验证

### 3.1 试验环境

为对提出的SUA-SAC算法及对比算法的性能

进行有效评估,所有试验均在统一的仿真平台上进行。试验所依赖的软硬件平台与关键环境参数如下。

硬件环境:所有模型的训练与测试均在同一台配置为Intel® Core™ i7-13700H CPU与NVIDIA GeForce RTX 4060 Laptop GPU的计算机上完成。

软件环境:基于Python 3.9及PyBullet物理引擎构建,算法实现基于PyTorch 1.13及stable-baselines3库。

仿真环境中的关键物理与环境参数设置见表1。

表1 仿真环境关键参数

Table 1 Key parameters of the simulation environment

参数	值	参数	值
无人机质量/kg	0.47	物理仿真时间步长/s	1/240
训练总步数	$8.0 \times 10^6$	经验回放缓冲区大小	$1.0 \times 10^6$
采样批量大小	256	折扣因子	0.99
网络学习率	0.0003	软更新系数	0.005
失败判定阈值/m	10.0	目标熵	-4
最小转弯半径/m	2.5	飞行空域边界/m	[50,50,30]
最大设计平飞速度/( $\text{m} \cdot \text{s}^{-1}$ )	15	最大设计抗风能力/( $\text{m} \cdot \text{s}^{-1}$ )	12

为确保所有对比试验均在一致的条件下进行,仿真试验所采用的各算法关键超参数见表2。

为了确保对比试验的公平性与有效性,表2中各算法的超参数取值均基于主流强化学习框架的推荐配置,并结合本文的任务特性进行了独立的预试验微调。特别是在学习率的设置上,考虑到PPO算法对策略更新步长具有较高的敏感性,过大的学习率极易导致策略性能崩溃且难以恢复,因此将其设定为较保守的 $1.0 \times 10^{-4}$ 以确保收敛稳定性;而对于SAC算法和双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3)算法,前者具备熵正则化机制,后者采用双延迟更新策略,两者在连续控制任务中均表现出较强的参数适应性,因此统一采用 $3.0 \times 10^{-4}$ 以在保证稳定的前提下提升收敛速度。所有算法均在最优的参数配置下进行评估,以反映其真实性能差异。

除了强化学习超参数外,UKF参数设定对状态估计性能至关重要。基于无人机动力学模型的特性与传感器标定数据,对UKF的关键参数进行了如下设定。

(1)初始状态协方差矩阵 $\xi_0$ :初始化为单位矩

表 2 关键超参数设置

Table 2 Key hyperparameter settings

超参数	SUA-SAC 算法	SAC 算法	TD3 算法	PPO 算法	描述
学习率/ $10^{-4}$	3.0	3.0	3.0	1.0	控制网络权重更新幅度
折扣因子	0.99	0.99	0.99	0.99	对未来奖励的重视程度
批处理大小	256	256	256		每次梯度更新时使用的经验样本数量
经验回放池大小/ $10^6$	1.0	1.0	1.0		存储历史经验以供重复学习的内存容量
目标熵	-4	-4			自动温度调整机制所要维持的策略熵
目标网络软更新系数	0.005	0.005	0.005		目标网络向网络软更新的混合比例
更新前步数				2 048	策略更新前,每个环境收集的经验步数
广义优势估计因子				0.95	用于平衡方差与偏差
更新轮次				10	每次收集数据后,对数据重复学习的次数
PPO 裁剪系数				0.2	PPO 算法中限制策略更新幅度的裁剪系数

阵  $0.1\mathbf{I}_{12 \times 12}$ , 表示初始时刻各状态量的估计方差均为 0.1。

(2) 过程噪声协方差矩阵  $\mathbf{\Gamma}$ : 其为过程噪声向量  $\mathbf{q}_k$  的协方差, 即  $\mathbf{q}_k \sim \mathcal{N}(0, \mathbf{\Gamma})$ 。设定  $\mathbf{\Gamma}$  为对角矩阵, 其中对应位置与速度状态的分量方差设为 0.01, 姿态分量方差设为 0.005。

(3) 基础测量噪声协方差矩阵  $\mathbf{G}_{\text{base}}$ : 位置测量噪声标准差设为  $\sigma_x = \sigma_y = 1.0 \text{ m}$ ,  $\sigma_z = 0.5 \text{ m}$ 。

为验证上述参数设定的合理性与鲁棒性, 本文进行了参数敏感性分析试验, 引入缩放因子  $\lambda$  对  $\mathbf{\Gamma}$  和  $\xi$  进行扰动测试, 测试结果如图 2 所示。

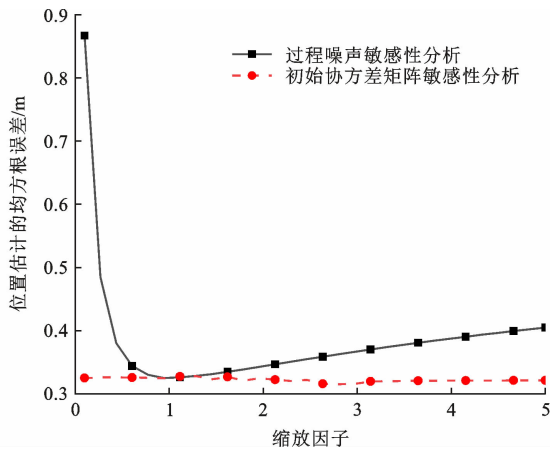


图 2 UKF 参数敏感性分析

Fig. 2 UKF parameter sensitivity analysis

图 2 中可看出初始协方差敏感性曲线近似水平, 表明算法对初始值的依赖度较低, 具备快速收敛能力。当缩放因子接近 1.0 时过程噪声敏感性曲线位置估计误差达到极小值; 当  $\lambda < 0.5$  时, 过小的过程噪声会导致滤波器过度信任模型而发散, 过大的噪声则引入不必要的抖动。该结果证实了模型选定

的  $Q$  值处于最优区间。对于受测量噪声影响的部分, 则采用第 1.2.2 节中提出的自适应测量噪声协方差矩阵  $\mathbf{G}_k$  的在线调整机制, 对基础协方差矩阵  $\mathbf{G}_{\text{base}}$  进行动态缩放, 从而保证不同工况下的状态估计精度。

### 3.2 试验结果

为验证所提出的 SUA-SAC 算法的有效性, 本节将从训练过程和多场景测试 2 个维度展开详细的对比分析。在对比算法的选择上, 采取了多维度的基准对照策略: 首先, 选取标准 SAC 算法作为同源基准, 以验证本文提出的不确定性感知模块的改进效果; 其次, 引入 TD3 算法作为确定性策略的强基线, 旨在对比分析确定性策略与 SAC 架构下的随机策略在复杂环境探索能力上的差异; 最后, 选取了在无人机控制领域广泛应用的 PPO 算法作为同策略 (On-policy) 基准, 以评估本文方法在样本效率与收敛速度上的相对优势。

#### 3.2.1 训练过程分析

为评估 SUA-SAC 算法与 SAC、PPO、TD3 等 3 种基线算法在学习效率与收敛性能上的差异, 本节分析了各算法在 800 万个训练步长内的关键指标变化, 如图 3 所示。

首先, 图 3(a) 的平均回合长度曲线反映了智能体掌握基本生存策略的速度。可以观察到, PPO 算法和标准 SAC 算法的回合长度率先大幅攀升并迅速稳定在最大设定步数, 表明这 2 类算法在低维状态输入下能够快速学会避免碰撞和越界的生存法则。相比之下, SUA-SAC 算法的回合长度收敛速度略慢于 PPO 算法和 SAC 算法。这是因为 SUA-SAC 算法需要处理包含状态协方差的高维输入, 并

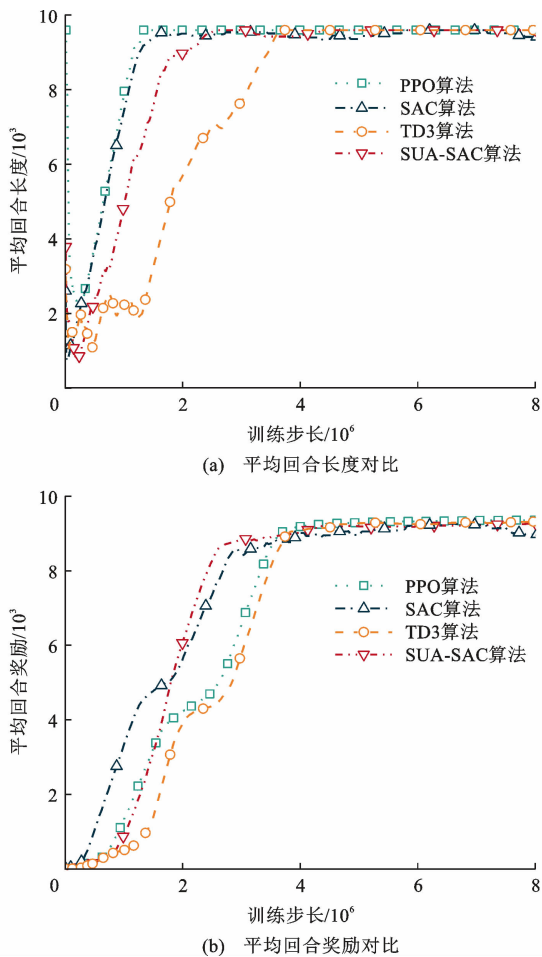


图3 四种算法学习曲线对比

Fig. 3 Comparison of learning curves for four algorithms

在策略中建立对不确定性的认知,这一复杂的特征提取过程在初期增加了学习难度。TD3算法则表现出最慢的收敛速度和明显的振荡,说明其确定性策略在探索复杂环境边界时存在局限性。

然而,尽管生存策略的形成稍晚,SUA-SAC算法在任务完成质量上却展现出显著优势。从图3(b)的平均回合奖励曲线可以观察到,SUA-SAC算法的奖励攀升速度更快,在约280万步左右即与SAC算法一同进入性能平台区,并最终收敛至最高的奖励水平。这表明SUA-SAC算法学习了基础飞行能力,其不确定性感知机制能够迅速引导智能体优化轨迹跟踪精度与控制平滑度。相比之下,PPO算法虽然生存能力形成较快,但其最终奖励值在约370万步处收敛。这是由于TD3算法缺乏有效的最大熵探索,在奖励获取上明显滞后,且在训练中后期仍存在波动。

SUA-SAC算法与标准SAC算法相对于PPO及TD3算法的优越性主要体现在图3(b)的奖励曲线对比中。由于PPO算法采用On-policy更新机

制,其样本利用率较低;而TD3算法受限于确定性策略的探索瓶颈,其收敛速度和最终稳定性均不及基于随机策略的算法SAC算法架构;相比之下,SUA-SAC算法在收敛速度上显著领先于所有基线算法。这说明面向数据获取成本高昂的无人机控制任务,融合不确定性感知的SAC算法展现出最优的学习速度。综上所述,提出的SUA-SAC算法不仅在最终性能上表现优异,更关键的是,其不确定性感知机制能够提升算法在复杂任务中的学习效率和稳定性。

### 3.2.2 多场景性能对比分析

设置了3种不同场景,定性的分析了各算法在三维空间环境下的控制效果,图4展示了不同算法在3种测试场景下的轨迹跟踪对比。

(1)场景1(静态无干扰基准场景):旨在测试算法在复杂林间地貌下的基础跟踪精度。规划路径模拟真实的林区公路形态,道路起点坐标为(0,25,5)m,终点坐标为(50,17,5)m,环境在道路两侧设置了静态长方体树冠模型以限制可行飞行空间,且不设置外部动态干扰。

(2)场景2(动态鸟类遮挡场景):旨在测试算法在感知受限时的鲁棒性。在场景1的基础上,引入一个球形障碍物(半径为0.3m)模拟林间飞鸟。该障碍物设定为以 $3\text{ m}\cdot\text{s}^{-1}$ 的速度沿直线横向穿过无人机飞行路径,并在 $t=25\text{ s}$ 时与无人机轨迹发生空间交叉,形成单次瞬时视线遮挡(遮挡持续约1~2s),以模拟突发的动态干扰。

(3)场景3(强风环境干扰场景):评估算法的抗扰动能力。在场景1的基础上加入非定常风场。设置背景平均风速为 $8\text{ m}\cdot\text{s}^{-1}$ ,风向与无人机初始航向呈 $90^\circ$ 垂直夹角,并叠加幅度为 $\pm 2\text{ m}\cdot\text{s}^{-1}$ 、频率为0.5 Hz的正弦强风扰动,以模拟林间穿堂风的湍流特性。

需要特别说明的是,鉴于本文聚焦于林区公路巡检任务,无人机的全局基准路径主要由道路中心线拓扑决定,因此在仿真中是预设的。

在场景1的静态无干扰环境中,各算法的性能差异已初步显现。PPO算法的轨迹在弯道处未能精确拟合目标曲线的几何形状,表现出较大的循迹偏差;新引入的TD3算法虽然大体沿路径飞行,但在曲线切入点和改出点呈现出一定的滞后与波动,控制精度略逊于标准SAC算法;标准SAC算法表现更优,但仍存在可感知的细微偏离;SUA-SAC算法则表现出最高的控制精度,其飞行轨迹与参考基

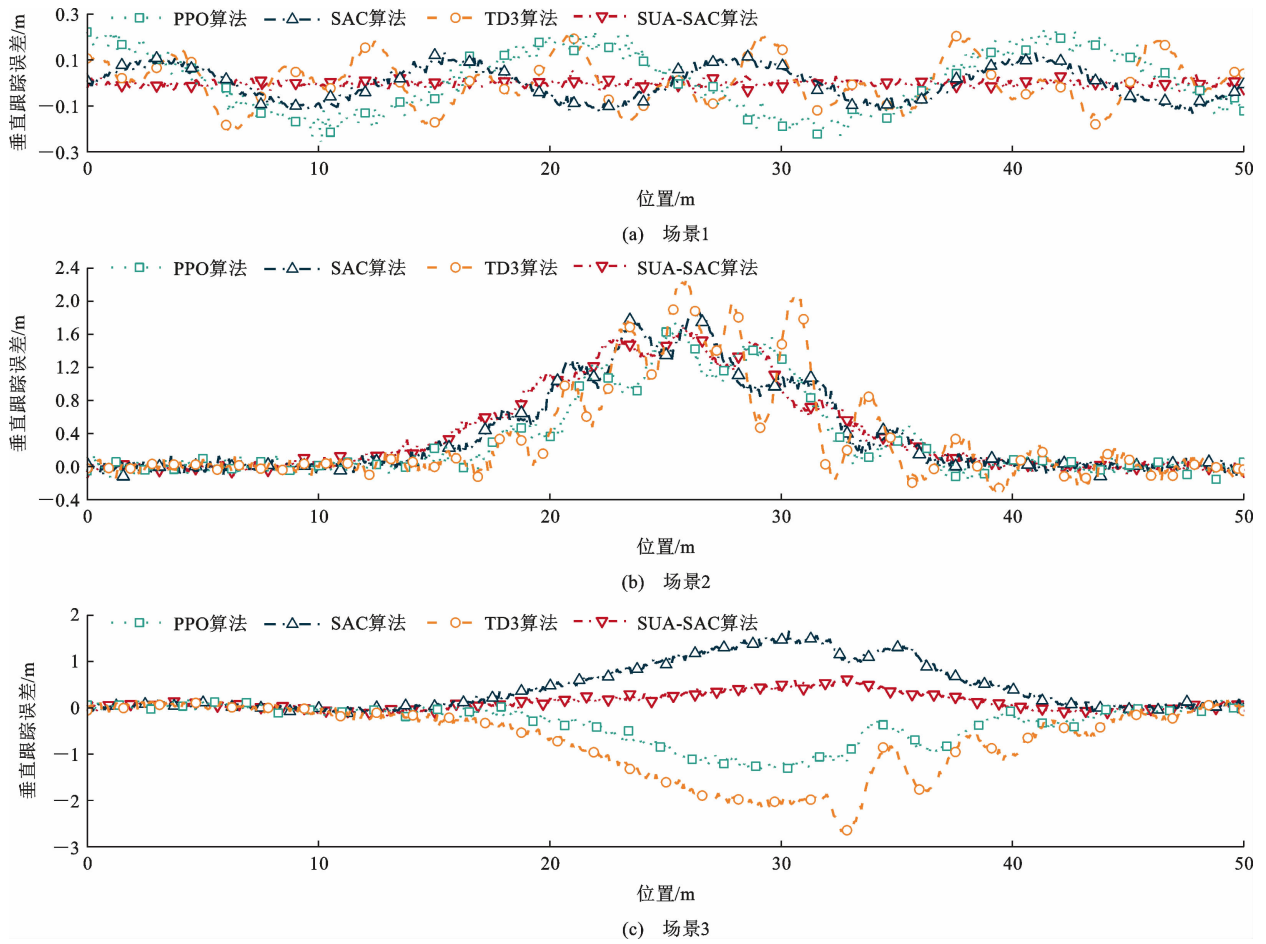


图 4 三种场景下算法跟踪路径效果对比

Fig. 4 Comparison of trajectory tracking performances of the algorithms under three scenario

准线保持了较高的一致性。这初步表明,通过不确定性感知进行训练所得到的策略本身具备了更精细化的控制能力。

从场景 2 的动态鸟类遮挡环境与场景 3 的强风干扰环境可以看出算法间的性能差距被放大。PPO 算法的轨迹发散较为严重,难以维持正常的巡航姿态。值得注意的是,TD3 算法在应对强外部干扰时,其轨迹呈现出显著的高频振荡特征,表明确定性策略在缺乏随机性探索辅助的情况下,难以在动态干扰中找到稳定的控制平衡点,容易产生过激的修正动作。标准 SAC 算法虽然得益于最大熵机制比 TD3 算法略显稳定,但其跟踪误差依然明显增大。相比之下,SUA-SAC 算法保持了较高的轨迹贴合度。这一现象为本文所提方法的核心机制提供了有力的视觉证据:当外部干扰发生时,UKF 的状态估计不确定性会急剧升高。标准 SAC、PPO 及 TD3 算法的策略网络以可能存在偏差的状态估计值为输入,因而易于产生过度或错误的控制修正,导致路径振荡甚至发散;而 SUA-SAC 算法的策略网络同时

以状态估计值及其不确定性为输入,在训练中可以在高不确定性区域执行更为平滑和保守的控制策略,有效抑制了干扰带来的不利影响。

在上述定性分析的基础上,为了精确量化各算法的轨迹跟踪性能,采用以下指标进行误差计算。首先,根据文中定义的坐标系与状态变量,定义  $k$  时刻的瞬时三维位置误差  $e_{p,k}$  为无人机在惯性系下离散后的实际位置向量  $\mathbf{x}_{E,k}$  与  $\mathbf{x}_{p,r,k}$  之间的欧氏距离,即

$$e_{p,k} = \|\mathbf{x}_{E,k} - \mathbf{x}_{p,r,k}\|_2 \quad (22)$$

基于此,平均跟踪误差 (ATE)、均方根误差 (RMSE) 以及最大误差分别为

$$E_{ATE} = \frac{1}{\chi} \sum_{k=1}^{\chi} e_{p,k} \quad (23)$$

$$E_{RMSE} = \sqrt{\frac{1}{\chi} \sum_{k=1}^{\chi} e_{p,k}^2} \quad (24)$$

$$E_{MaxError} = \max_{k=1}^{\chi} e_{p,k} \quad (25)$$

式中:  $\chi$  为测试回合的总时间步数。

表 3 详细列出了 SUA-SAC、SAC、TD3 及 PPO 算法在 3 种测试场景下的关键性能指标。数据分析

表3 各算法在不同场景的性能指标

Table 3 Performance metrics of different algorithms in various scenarios

性能指标	场景 1				场景 2				场景 3			
	PPO	SAC	TD3	SUA-SAC	PPO	SAC	TD3	SUA-SAC	PPO	SAC	TD3	SUA-SAC
任务成功率/%	98.1	99.1	97.9	<b>99.7</b>	95.2	98.6	94.8	<b>99.8</b>	94.5	94.8	93.2	<b>98.2</b>
平均跟踪误差/m	1.76	0.85	1.82	<b>0.28</b>	2.31	1.15	2.85	<b>0.45</b>	2.95	1.48	3.42	<b>0.51</b>
均方根误差/m	2.05	1.02	2.25	<b>0.34</b>	2.85	1.38	3.45	<b>0.53</b>	3.51	1.75	4.05	<b>0.62</b>
最大误差/m	2.64	1.28	2.98	<b>0.41</b>	3.24	1.95	4.85	<b>0.88</b>	4.18	2.65	5.25	<b>1.15</b>
误差标准差/m	0.95	0.45	1.12	<b>0.19</b>	1.22	0.61	1.65	<b>0.25</b>	1.55	0.78	1.95	<b>0.31</b>
路径长度比	1.012	1.005	1.015	<b>1.001</b>	1.021	1.009	1.032	<b>1.004</b>	1.028	1.013	1.048	<b>1.006</b>
累计绝对急动度	18.2	9.8	19.7	<b>4.5</b>	29.8	15.5	38.5	<b>8.1</b>	35.5	19.2	48.6	<b>9.5</b>
控制能耗	0.82	0.65	0.87	<b>0.48</b>	0.91	0.72	1.12	<b>0.55</b>	0.98	0.81	1.35	<b>0.61</b>
单步平均求解 耗时/ms	0.82	0.88	0.82	<b>1.56</b>	0.81	0.89	0.81	<b>1.58</b>	0.83	0.88	0.83	<b>1.55</b>

将围绕精确性、鲁棒性与安全性、平顺性与效率 3 个核心维度展开。

在精确性方面,在无干扰的场景 1 中,SUA-SAC 算法的 ATE 相较于 SAC 算法下降了 67%,相较于 PPO 算法和 TD3 算法下降了 80%以上。值得注意的是,TD3 算法作为确定性策略算法,在缺乏最大熵机制辅助探索的情况下,策略容易陷入次优解,导致基础控制精度不足。

在鲁棒性与安全性方面,从任务成功率来看,SUA-SAC 算法在场景 3 中仍能达到 98.2%的成功率。相比之下,TD3 算法在强风干扰下的成功率较低,且最大瞬时误差较大,标明了标准确定性策略对环境扰动极为敏感,容易导致控制发散;而 SUA-SAC 算法通过在高不确定性时采取保守策略,有效避免了极端误差导致的控制失效。

最后,在平顺性与效率方面,累计绝对急动度和控制能耗指标揭示了飞行品质的巨大差异。TD3 算法在所有场景下均记录了较高的急动度,呈现出控制振荡特征。相比之下,SUA-SAC 算法的累计急动度最小,意味着其飞行轨迹最为平滑,避免了不必要的急加速和急转弯。更平滑的控制直接带来了更低的控制能耗,从而能有效延长无人机的续航时间和作业半径。

此外,为了评估算法的工程适用性,本文在试验硬件平台上对各算法的计算效率进行了测试,并统计了单步控制循环中的平均求解耗时。结果显示,SUA-SAC 算法的单步平均推理延迟约为 1.56 ms。虽然相较于基线算法增加了因 UKF 滤波带来的计算开销,但该耗时仍显著低于仿真环境 240 Hz 的物

理控制周期。这表明本文方法仅占用了约 37%的时间片资源,留有充足的计算余量用于通信与系统监控,具备在机载嵌入式平台上进行高频实时控制的可行性。

### 3.3 消融试验

消融试验将本文提出的完整方法(SUA-SAC 算法+自适应 UKF)与移除部分创新点的变体进行性能对比(图 5),具体包括:消融 A(标准 SAC+自适应 UKF,移除了不确定性感知机制)和消融 B(SUA-SAC 算法+标准 UKF,移除了自适应噪声模型),并以标准 SAC+标准 UKF 作为统一的基准方法。

首先,从图 5(a)的消融试验学习曲线可以看出各组件对训练效率的综合影响。本文方法在收敛速度和最终奖励上表现优异。移除自适应噪声模型的消融 B 曲线与基准方法的学习曲线均显著劣于具备自适应滤波的 2 种方法。这一现象表明,在存在显著过程噪声和测量噪声的环境中,鲁棒的状态估计器是强化学习能够成功收敛的基础。若状态估计本身存在巨大偏差,即使策略网络具备不确定性感知能力,也会导致学习效率低下。

为进一步辨析各组件在最终控制性能上的作用,本文对各训练完成的模型进行了轨迹跟踪误差分析。最后,选取了最具代表性与挑战性的场景 3 进行对比,因为其持续且不可预测的外部力干扰,能最大程度地激发状态估计的不确定性,从而最清晰地揭示不同算法在鲁棒性上的差异。

由图 5(b)可知,本文方法的误差在较低水平,消融 A 得益于前端融合带来的高精度滤波,在 30 m

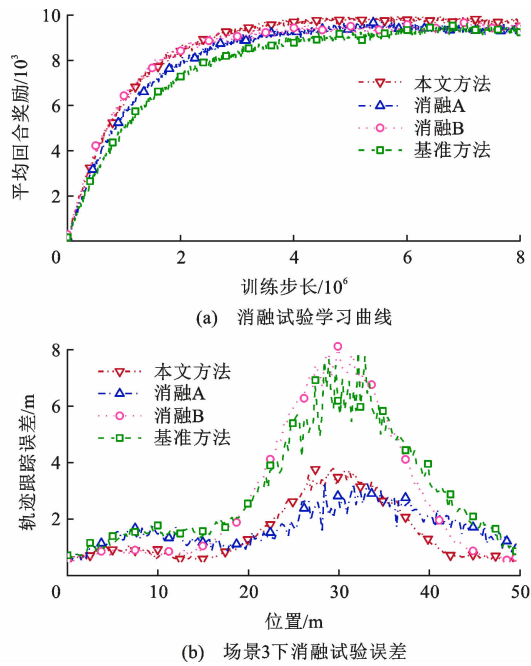


图 5 算法消融试验对比

Fig. 5 Comparison of algorithm ablation experiments

路段展现出较低的瞬时误差。但是,纵观全程,由于标准 SAC 算法策略网络忽略了融合系统输出的残余不确定性,导致其在应对干扰时采取了过于激进的修正动作,呈现出明显的高频振荡特征。相比之下,本文方法将不确定性信息显式融入决策回路,虽然在个别极值点上略显保守,但整体轨迹更为平滑,

方差更小。这说明单纯依靠前端的状态估计融合尚不足以解决复杂环境下的鲁棒控制问题,必须结合后端的不确定性感知机制,才能在精度与稳定性之间取得最佳平衡。

### 3.4 GPS 信号遮挡下误差分析

无人机在真实环境中飞行时,由于森林树木密集,遮挡众多,不可避免地会出现 GPS 信号丢失的情况。因此,在仿真测试中,首先让智能体执行 1 次 60 s 的飞行任务,并在第 20~40 s 期间,人为中断 GPS 信号的输入,以模拟其在森林树冠下飞行的真实场景,图 6 记录了仿真测试过程中各算法的横向与高程跟踪误差。

SUA-SAC 算法的误差曲线在 GPS 信号丢失区域内,表现出了较为稳定的性能,从横向相对误差与高程相对误差 2 个方面展现了鲁棒性。该试验过程实质上也是对不确定性影响决策机制的一次动态可视化验证:在 20~40 s 内,随着 GPS 信号中断,输入状态中的协方差分量急剧增加。从图 6 中 TD3 算法的误差发散可以看出,若策略忽略这一不确定性变化,盲目信任惯性漂移将导致控制失效。而 SUA-SAC 算法平稳的轨迹则证明了算法成功识别到了这一高不确定性特征,并动态调整了控制策略,避免了激进的修正动作。这证实了算法具备根据协方差信息实时切换策略的能力。

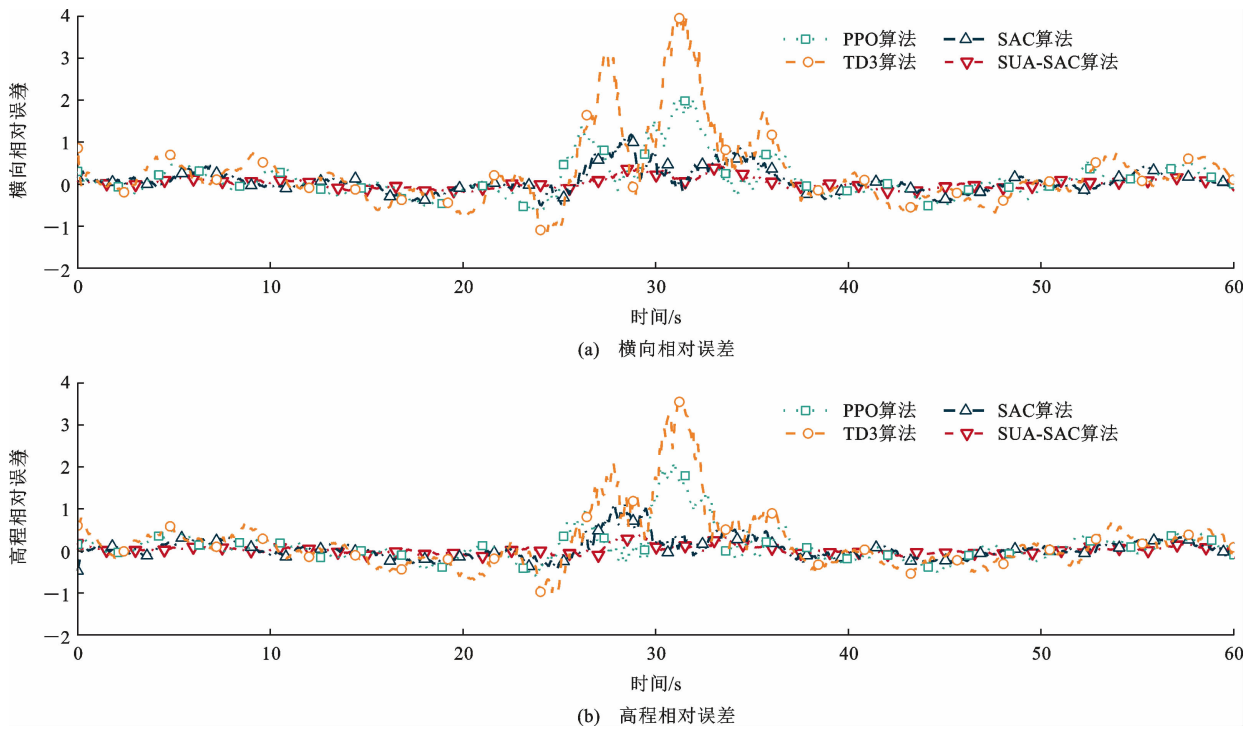


图 6 GPS 信号遮挡下横向与高程相对误差分析

Fig. 6 Analysis of lateral and elevation relative errors under GPS signal obstruction

虽然本试验主要展示了 GPS 信号完全中断,即可靠性因子  $\lambda_k \rightarrow 0$  时的极端工况,但 AMNC 机制在理论上同样适用于多径效应、信号漂移等复杂扰动模式。在真实林区环境中,当 GPS 信号受到树冠部分遮挡或发生多径反射时,信号质量通常表现为信噪比的连续下降,在这种情况下,自适应 UKF 将依据  $\lambda_k$  执行更平滑的权重调整。

综上,通过验证算法在信号完全中断这一最恶劣边界条件下的稳定性,实际上也确立了其在各类非中断型信号扰动下的性能下界,充分证明了该框架在真实复杂电磁环境下的有效性。这对于保障无人机在信号遮挡频繁的真实复杂环境中执行任务时的安全性与可靠性至关重要。

## 4 结 语

(1)本文系统地研究了无人机在林区公路感知受限环境下的高精度轨迹跟踪深度强化学习控制问题。通过设计一种自适应无迹卡尔曼滤波算法,解决了 GPS 信号缺失时的鲁棒状态估计问题,并将量化后的估计不确定性信息作为关键要素输出。提出了 SUA-SAC 算法,通过将不确定性信息融入决策网络,使智能体学会了在高置信度下追求精度、在低置信度下保障安全的智能控制策略。仿真试验从学习效率、跟踪精度、鲁棒性和飞行效率等多个维度,验证了模型的优越性。

(2)本文最重要的实用价值在于,为无人机在 GPS 信号受扰或缺失的环境下的自主飞行,提供了一种行之有效的解决方案。所提方法能够显著提升无人机在恶劣环境下的任务成功率和飞行安全性,避免因感知信息退化而导致的控制失效。

(3)未来应考虑将模型部署到真实的无人机硬件平台。研究应重点引入域随机化技术,在训练中对物理参数进行宽范围的随机扰动,迫使智能体学习适应模型参数的不确定性分布;同时,结合系统辨识技术对真实无人机的动力学参数进行精细标定,以进一步弥合仿真与现实的差距,提升控制策略在真实林区环境下的鲁棒性。

### 参 考 文 献 :

#### References :

[1] 杨 豪,刘李彦,张军辉,等. 环境适应性优化的轻量化多尺度道路裂缝检测[J]. 中国公路学报,2025,38(7):118-134.  
YANG Hao, LIU Li-yan, ZHANG Jun-hui, et al. Environmental adaptability optimization lightweight multi-scale road crack detection[J]. China Journal of Highway and

Transport, 2025, 38(7): 118-134.

[2] SILVA L A, LEITHARDT V R Q, BATISTA V F L, et al. Automated road damage detection using UAV images and deep learning techniques[J]. IEEE Access, 2023, 11: 62918-62931.

[3] TUTSOY O, ASADI D, AHMADI K, et al. Minimum distance and minimum time optimal path planning with bioinspired machine learning algorithms for faulty unmanned air vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(8): 9069-9077.

[4] LOPEZ-SANCHEZ I, MORENO-VALENZUELA J. PID control of quadrotor UAVs: A survey[J]. Annual Reviews in Control, 2023, 56: 100900.

[5] MOHINDRU P. Review on PID, fuzzy and hybrid fuzzy PID controllers for controlling non-linear dynamic behaviour of chemical plants[J]. Artificial Intelligence Review, 2024, 57(4): 97.

[6] ZHAO B, XIAN B, ZHANG Y, et al. Nonlinear robust adaptive tracking control of a quadrotor UAV via immersion and invariance methodology [J]. IEEE Transactions on Industrial Electronics, 2015, 62(5): 2891-2902.

[7] 申富媛,李 炜,蒋栋年. 四旋翼无人机寿命预测和自主维护方法[J]. 吉林大学学报(工学版),2023,53(3):841-852.  
SHEN Fu-yuan, LI Wei, JIANG Dong-nian. Life prediction and self-maintenance method of quadrotor unmanned aerial vehicle[J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(3): 841-852.

[8] 秦明星,王 忠,李海龙,等. 基于分布式模型预测的无人机编队避障控制[J]. 北京航空航天大学学报,2024,50(6):1969-1981.  
QIN Ming-xing, WANG Zhong, LI Hai-long, et al. Obstacle avoidance control of UAV formation based on distributed model prediction [J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(6): 1969-1981.

[9] XIE R L, MENG Z J, WANG L F, et al. Unmanned aerial vehicle path planning algorithm based on deep reinforcement learning in large-scale and dynamic environments[J]. IEEE Access, 2021, 9: 24884-24900.

[10] YIN Y F, GUO Y, SU Q R, et al. Task allocation of multiple unmanned aerial vehicles based on deep transfer reinforcement learning[J]. Drones, 2022, 6(8): 215.

[11] ZUO Z Y, LIU C J, HAN Q L, et al. Unmanned aerial vehicles: Control methods and future challenges[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(4): 601-614.

[12] BOUHAMED O, GHAZZAI H, BESBES H, et al. Autonomous UAV navigation: A DDPG-based deep reinforcement learning approach [C] // IEEE. 2020 IEEE International Symposium on Circuits and Systems (ISCAS). New York: IEEE, 2020: 1-5.

[13] 郭 靖,鲜 勇,任乐亮,等. 自适应斥力势场的无人机航迹规划[J/OL]. 北京航空航天大学学报,2024,https://doi.org/10.13700/j. bh. 1001-5965. 2024. 0569.

- GUO Jing, XIAN Yong, REN Le-liang, et al. Adaptive Repulsive Potential Field for UAV Trajectory Planning[J/OL]. *Journal of Beijing University of Aeronautics and Astronautics*, 2024, <https://doi.org/10.13700/j.bh.1001-5965.2024.0569>.
- [14] ZHAN G, ZHANG X M, LI Z C, et al. Multiple-UAV reinforcement learning algorithm based on improved PPO in ray framework[J]. *Drones*, 2022, 6(7): 166.
- [15] 陈运翔, 苟明, 张建平, 等. 基于多智能体近端策略优化的低空异构飞行器实时三维冲突解脱方法[J]. *交通运输工程学报*, 2026, 26(3): 185-197.
- CHEN Yun-xiang, GOU Ming, ZHANG Jian-ping, et al. Real-time 3D conflict resolution method for low-altitude heterogeneous aircraft based on multi-agent proximal policy optimization[J]. *Journal of Traffic and Transportation Engineering*, 2026, 26(3): 185-197.
- [16] MA B D, LIU Z B, DANG Q Q, et al. Deep reinforcement learning of UAV tracking control under wind disturbances environments[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1-13.
- [17] USMAN M, IRSHAD A, ASHRAF CHAUDHRY S. SAC-DeV: Secure access control for drone-assisted edge computing in e-VANETs[J]. *IEEE Internet of Things Journal*, 2025, 12(24): 53121-53131.
- [18] 寇凯, 杨刚, 张文启, 等. 基于SAC的无人机自主导航方法研究[J]. *西北工业大学学报*, 2024, 42(2): 310-318.
- KOU Kai, YANG Gang, ZHANG Wen-qi, et al. Exploring UAV autonomous navigation algorithm based on soft actor-critic[J]. *Journal of Northwestern Polytechnical University*, 2024, 42(2): 310-318.
- [19] 雷耀麟, 丁文锐, 罗祎喆, 等. 无人机数据采集任务中的航迹规划与资源分配优化[J]. *北京航空航天大学学报*, 2025, 51(10): 3460-3470.
- LEI Yao-lin, DING Wen-rui, LUO Yi-zhe, et al. Trajectory planning and resource allocation optimization in UAV data collection missions[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2025, 51(10): 3460-3470.
- [20] 赵军, 何家政, 孙冰寒, 等. 基于深度强化学习的四旋翼无人机姿态控制[J]. *中国惯性技术学报*, 2025, 33(3): 284-292, 300.
- ZHAO Jun, HE Jia-zheng, SUN Bing-han, et al. Attitude control of quadrotor unmanned aerial vehicle based on deep reinforcement learning[J]. *Journal of Chinese Inertial Technology*, 2025, 33(3): 284-292, 300.
- [21] MITTAL V, MAGHSUDI S, HOSSAIN E. Distributed cooperation under uncertainty in drone-based wireless networks: A Bayesian coalitional game[J]. *IEEE Transactions on Mobile Computing*, 2023, 22(1): 206-221.
- [22] 黄鹤, 张科, 陈永安, 等. 一种无人机航拍目标的长期跟踪算法[J]. *哈尔滨工业大学学报*, 2022, 54(5): 104-116.
- HUANG He, ZHANG Ke, CHEN Yong-an, et al. A long-term tracking algorithm for UAV aerial photography[J]. *Journal of Harbin Institute of Technology*, 2022, 54(5): 104-116.
- [23] 张宏展, 韩鹏, 赵可, 等. 融合金字塔栅格与测向交叉的低空无人机定位[J]. *交通信息与安全*, 2025, 43(3): 162-170.
- ZHANG Hong-zhan, HAN Peng, ZHAO Ke, et al. Low-altitude UAV positioning fusing pyramid grid and direction-finding cross-location[J]. *Journal of Transport Information and Safety*, 2025, 43(3): 162-170.
- [24] 侯佳林, 侯榕榕, 鲍跃全. 基于三维高斯溅射的桥梁自动化三维重建方法[J]. *中国公路学报*, 2026, 39(1): 87-98.
- HOU Jia-lin, HOU Rong-rong, BAO Yue-quan. Automated 3D reconstruction of bridges based on 3D Gaussian splatting[J]. *China Journal of Highway and Transport*, 2026, 39(1): 87-98.
- [25] 李书恒, 何德峰, 廖飞, 等. 基于非线性前馈补偿的六旋翼无人机姿态稳定反步控制[J]. *上海交通大学学报*, 2025, 59(12): 1891-1900.
- LI Shu-heng, HE De-feng, LIAO Fei, et al. Backstepping control for attitude stabilization of hexacopter based on nonlinear feedforward compensation[J]. *Journal of Shanghai Jiao Tong University*, 2025, 59(12): 1891-1900.
- [26] 房杰, 谢富源, 曹越, 等. 一种四旋翼无人机的三维轨迹规划与抗干扰轨迹跟踪控制方法[J]. *南京理工大学学报*, 2025, 49(4): 494-503.
- FANG Jie, XIE Fu-yuan, CAO Yue, et al. A three-dimensional trajectory planning and anti-disturbance trajectory tracking control method of quadcopters[J]. *Journal of Nanjing University of Science and Technology*, 2025, 49(4): 494-503.
- [27] 蔡志浩, 陈文军, 赵江, 等. 基于动态视觉传感器的无人机目标检测与避障[J]. *北京航空航天大学学报*, 2024, 50(1): 144-153.
- CAI Zhi-hao, CHEN Wen-jun, ZHAO Jiang, et al. Object detection and obstacle avoidance based on dynamic vision sensor for UAV[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2024, 50(1): 144-153.
- [28] 羊钊, 齐洪彪, 于阳阳, 等. 风险规避与组合策略融合的多无人机协同路径规划方法[J]. *交通运输工程学报*, 2026, 26(3): 140-158.
- YANG Zhao, QI Hong-biao, YU Yang-yang, et al. Integrated risk avoidance and hybrid strategy for multi-UAV cooperative path planning[J]. *Journal of Traffic and Transportation Engineering*, 2026, 26(3): 140-158.