

基于 PPO – SAC 的杂交水稻制种授粉机组 非规则路径协同控制研究

王永维¹ 史荣凯¹ 李英² 奚特¹ 冯槐区¹ 王俊¹

(1. 浙江大学生物系统工程与食品科学学院, 杭州 310058; 2. 上海联适导航技术股份有限公司, 上海 201702)

摘要: 授粉是杂交水稻制种的关键环节,针对杂交水稻制种机械拉绳碰击式授粉在非规则田块中难以实现精准位速协同控制的问题,提出了一种基于近端策略优化(Proximal policy optimization,PPO)和软性演员评论家(Soft actor-critic,SAC)的双环强化学习机组协同控制算法。将授粉车辆的运动控制解耦为具有相对独立目标的子系统,横向上将路径曲率及其变化率融合到 PPO 算法框架实现对田间弯曲路径的精确跟踪,纵向上将机组成员相对纵向位置与输出转向角融合到 SAC 算法框架实现机组车辆位速控制,构成双环 PPO – SAC 强化学习算法实现授粉车辆的精准协同,通过仿真试验与模拟验证试验验证算法的有效性。双环 PPO – SAC 算法表现优于单框架的强化学习算法,其平均路径跟踪偏差为 0.161 m,相比于单 PPO 算法降低 29.9%,机组纵向相对距离偏差相比于单 SAC 算法减少 22.5%,算法验证试验中机组纵向相对距离偏差控制在 0.325 m 内,田间实地作业平均协同偏差为 0.585 m,满足作业要求。本文研究提供了一套适用于杂交水稻制种授粉的非规则路径协同控制算法,该算法具有精准的路径跟踪性能和机组协同性能,为杂交水稻制种的现代化和自动化作业奠定了基础。

关键词: 杂交水稻授粉机组; 强化学习; 协同控制; PPO – SAC

中图分类号: S24

文献标识码: A

文章编号: 1000-1298(2026)19-0215-11

OSID:



Collaborative Control of Irregular Paths for Hybrid Rice Seed Production Pollination Units Based on PPO – SAC

Wang Yongwei¹ Shi Rongkai¹ Li Ying² Xi Te¹ Feng Huaigu¹ Wang Jun¹

(1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

2. Shanghai AllyNav Technology Co., Ltd., Shanghai 201702, China)

Abstract: Pollination is a crucial link in hybrid rice seed production. Aiming at the problem that precise position-velocity collaborative control is difficult to achieve in irregular fields for mechanical rope-impact pollination in hybrid rice seed production, a dual-loop reinforcement learning-based collaborative control algorithm for pollination units was proposed, which integrated proximal policy optimization (PPO) and soft actor-critic (SAC). The motion control of pollination vehicles was decoupled into subsystems with relatively independent objectives: in the lateral direction, path curvature and its rate of change were integrated into the PPO algorithm framework to achieve accurate tracking of curved field paths; in the longitudinal direction, the relative longitudinal positions of unit members and output steering angles were fused into the SAC algorithm framework to realize position-velocity control of master-slave vehicles. A dual-loop PPO – SAC reinforcement-learning algorithm was constructed to realize precise coordination of pollination vehicles. The effectiveness of the algorithm was verified through simulation tests and field experiments. The results demonstrated that the dual-loop PPO – SAC algorithm outperformed single-framework reinforcement learning algorithms. Its average path tracking error reached 0.161 m, which was 29.9% lower than that of the single PPO algorithm; the longitudinal relative distance error was reduced by 22.5% compared with that of the single SAC algorithm. In the simulated field operation test, the longitudinal relative distance error was controlled within 0.325 m, and the average coordination error in actual field operation was 0.585 m, which met the operational requirements. The research result can

收稿日期: 2025 – 12 – 25 修回日期: 2026 – 04 – 24

基金项目: 国家重点研发计划项目(2023YFD2000403)和浙江省“三农九方”科技协作计划项目(2023SNJF048)

作者简介: 王永维(1975—),男,教授,博士,主要从事杂交水稻制种授粉机械研究,E-mail: wywzju@zju.edu.cn

provide a set of irregular path collaborative control algorithms suitable for hybrid rice seed production pollination, featuring precise path tracking performance and collaborative operation capabilities, which laied a foundation for the modernization and automation of hybrid rice seed production.

Key words: hybrid rice pollination units; reinforcement learning; collaborative control; PPO – SAC

0 引言

杂交水稻作为全球主要的粮食作物之一^[1-3], 在世界粮食安全中发挥着重要作用。授粉是杂交水稻种子生产的重要环节^[4], 进行充分均匀的授粉是提高杂交水稻种子产量与品质的重要途径^[5-6]。但水稻作为非严格自花授粉作物, 其自然杂交率仅有 0.2% ~ 0.3%, 单穗开花期仅有 5 ~ 7 d, 主要集中在每日 09:00—11:00, 有效授粉窗口窄、时效性要求严苛。因此, 开展高效、精准的人工辅助授粉作业, 成为提升水稻授粉率和结实率的关键。

人工辅助授粉主要包括人力授粉和机械授粉。人力授粉通过单长竿、双短竿碰撞父本穗部实现授粉, 劳动强度大、授粉效率低。机械授粉包含气力式和碰击式授粉。气力式授粉通过风机产生辅助气流将父本花粉吹到母本上^[7-9], 可实现较高质量的授粉, 但在授粉过程中需沿着父本行反复作业, 仍存在授粉效率低的问题; 无人机授粉通过飞行产生的气流扰动父本茎秆完成花粉的传播^[4], 可适应复杂山区地形作业, 但其花粉的扩散轨迹复杂且在大规模制种田中仍存在授粉效率不高的问题^[10]。碰击式授粉通过绳索模拟大规模人力授粉, 利用惯性和自然风将花粉传播到母本柱头上^[5,11]。基于这一原理通过 2 台动力底盘牵引授粉绳索, 在目标田块两侧以设定的速度和位置实现高效授粉作业。Xi 等^[12]的研究表明授粉绳索碰击父本的角度、速度会影响到花粉脱落的效率与扩散轨迹, 且当前的制种田中存在大量天然的非规则田块, 人工驾驶难以实现授粉机组精准的速度和位置协同, 使得授粉作业不均匀, 影响母本授粉效果。

农机协同作业模式在收获-运粮、耕整地等领域已经得到了广泛应用^[13-16]。其中, 基于“角度-距离”和“距离-距离”的领航跟随法是目前使用最为广泛的方法之一^[17-18]。现有的研究多通过 PID 等传统控制算法, 以固定的角度和距离实现直线作业路线下的农机协同作业^[19-21]。但在杂交水稻机组协同授粉中, 2 台授粉车辆之间的距离不是固定的, 需要根据道路变化对速度和转向进行实时调控, 传统控制方法如模糊控制^[22]、PID^[23-24]、滑模控制^[25-27] (Sliding mode control, SMC) 等依赖于精确的数学模型, 无法实现复杂路径状态下的实时调节。

强化学习为农业复杂作业提供了灵活自主的学习机制^[28-30], 不依赖精确数学模型或者高质量数据, 即可通过奖惩机制与网络结构为农机提供实时精确控制指令^[31-32], 实现了在动态场景下的多障碍编队、协同规划等复杂控制难题^[33-35]。然而, 该方法在实际应用中仍面临样本效率低与超参数调优复杂等问题, 大大增加整体任务的执行难度。因此, 在杂交水稻制种授粉这种路径不规则、需实时动态协同的复杂场景中, 现有方法存在灵活性不足、环境适应性较差的问题, 难以满足精准协同授粉的作业要求。

为实现在非规则制种田授粉机组的精确速度控制与位置协同, 保障授粉效果, 本研究提出一种基于 PPO – SAC 的杂交水稻制种授粉非规则地块协同控制算法。该算法拟通过双环强化学习的训练思路, 将授粉任务解耦为横向路径跟踪任务和纵向协同任务, 分别采用融合路径信息的 PPO 算法和融合相对位置信息的 SAC 算法进行训练, 实现路径跟踪与机组协同功能的独立优化。基于 PPO – SAC 算法, 进一步开展杂交水稻制种授粉仿真试验与模拟验证试验, 系统验证其在实际作业环境下的作业性能。

1 协同控制算法设计

1.1 杂交水稻制种机组协同授粉作业模式

杂交水稻制种机组协同授粉由 2 台授粉车辆与授粉绳索组成 (图 1), 将绳索两端分别紧固在授粉车辆尾部, 2 台授粉车沿垂直于水稻种植行方向同步行驶, 牵引授粉绳索进行授粉作业。2 台授粉车辆均搭载控制器, 领航车沿指定作业路径进行路径跟踪, 并实时将自身位置与状态传输到跟随车, 跟随车计算得到期望的目标位置与速度并通过内置控制算法解算, 完成速度与位置的协同。授粉绳索以一定的速度碰击父本水稻穗部, 使花粉飘向目标并沉降至母本柱头, 完成授粉作业。

1.2 授粉车辆运动学模型

Frenet 坐标系是一种随路径曲线动态变化的局部坐标系, 通过道路切线方向弧长 s 和与切线垂直且法向指向道路内侧 (道路弯道的曲率中心所在的一侧) 的方向横向偏差 l 构成坐标系 (图 2), 能够直接以参考路径点为基准表征车辆的运行状态, 同时解耦纵横向运动, 有助于改善跟踪精度与协同控制效果。Frenet 坐标系下的授粉车辆运动学公式为

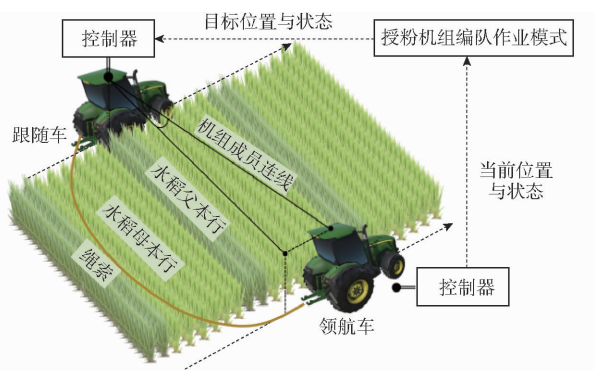


图 1 协同授粉作业模式

Fig. 1 Collaborative pollination operation mode

$$\begin{cases} \dot{s} = \frac{v \cos \theta}{1 - k_r l} \\ \dot{l} = v \sin \theta \\ \theta = \varphi - \varphi_r \\ \dot{\theta} = \frac{v \tan \delta}{L} - k_r \dot{s} \end{cases} \quad (1)$$

式中 $\dot{s}$ ——弧长变化率, m/s
 v ——授粉车辆作业速度, m/s
 θ ——航向角偏差, rad
 k_r ——参考轨迹的路径曲率
 $\dot{l}$ ——横向偏差变化率, m/s
 φ ——授粉车辆实际航向角, rad
 φ_r ——授粉车辆参考路径投影点切线角, rad
 $\dot{\theta}$ ——航向角偏差变化率, rad/s
 δ ——前轮转向角, rad
 L ——车辆轴距, m

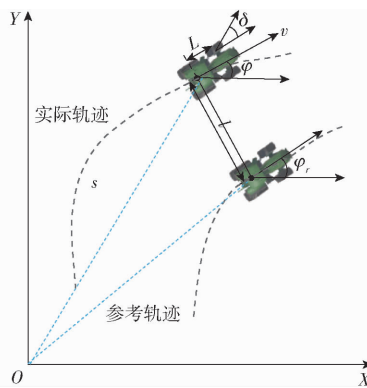


图 2 Frenet 下的授粉车辆运动学模型

Fig. 2 Kinematic model of pollination vehicle under Frenet

1.3 路径获取与预处理

作业前授粉车辆需预先采集田间作业路径,通过无人机以统一采集高度 5 m、采集速度 2 m/s,在道路中心线飞行采集路径点,采集完成后将信息上传到信息共享平台。但是直接采集的路径点易受田间卫星信号波动与地面干扰影响,通常不满足平滑性要求,需进行预处理。

设原始路径点集为 $\{P_i(x_i, y_i)\}_{i=1}^n$, 以路径点采集顺序排序, 再通过移动均值滤波去除噪点

$$\tilde{P}_i = \frac{1}{1 + 2k} \sum_{j=i-k}^{i+k} P_j \quad (2)$$

式中 $\tilde{P}_i$ ——经滤波平滑后的第 i 个路径点

k ——窗口尺寸

j ——路径点索引

为避免路径曲率突变(无回环、竖直路段),采用三次样条插值对路径进行平滑处理,在区间 $[x_i, x_{i+1}]$ 内,有三次多项式

$$y(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (3)$$

式中 a_i, b_i, c_i, d_i ——多项式系数

需要满足连续性条件

$$\begin{cases} y(x_i) = y_i \\ y(x_{i+1}) = y_{i+1} \\ y(x_{i+1}) = a_i + b_i h_i + c_i h_i^2 + d_i h_i^3 \\ h_i = x_{i+1} - x_i \\ y'(x) = b_i + 2c_i(x - x_i) + 3d_i(x - x_i)^2 \\ b_i = y'(x_i) = m_i \\ y''(x) = 2c_i + 6d_i(x - x_i) \\ y''_1(x_1) = 0 \\ y''_{n-1}(x_n) = 0 \end{cases} \quad (4)$$

式中 h_i ——第 i 个相邻路径点的横坐标差

m_i ——第 i 个路径点处的一阶导数值

为实现授粉车辆对路径状态的快速预测,引入未来一段距离路径的平均曲率变化率 κ 作为参考输入,计算公式为

$$\kappa = \frac{|y''(x)|}{[1 + (y'(x))^2]^{3/2}} \quad (5)$$

当车辆位于 $x = x_c$, x_c 为路径上随机一点,取未来距离 s_0 内的 $N + 1$ 个离散点,计算曲率序列 $\{\kappa_k\}_{k=0}^N$, 由于曲率变化率本质上是曲率对弧长的导数,通过离散差分近似可计算得出曲率的平均变化率 $\dot{\kappa}_{\text{avg}}$ 为

$$\dot{\kappa}_{\text{avg}} = \frac{\kappa_N - \kappa_0}{s_0} \quad (6)$$

式中 κ_0 ——当前位置的路径曲率

κ_N ——末端采样点的路径曲率

授粉车辆作业时,可通过预先采集的路径点与当前位置计算平均曲率变化率。

1.4 双环强化学习设计

1.4.1 网络架构

杂交水稻制种协同授粉中,授粉车辆需要兼顾路径精确跟踪与机组协同,单一网络的奖励权重分

配难度大、训练耗时长。现结合农业场景中车辆运动特性的独立性,将车辆运动纵横向解耦为路径跟踪与机组协同任务,采用双环分层学习框架:路径跟踪任务需保证算法训练的精度与稳定性,PPO 可通过 CLIP 裁剪操作稳步更新训练策略,实现稳步高精度的任务学习;机组协同任务需要能够快速应对不

同相对位置的变换,采用 SAC 通过最大熵函数实现对不同变化及动作的探索,契合机组协同的任务需求。如图 3 所示,阶段 I 通过 PPO 训练车辆的转向跟踪能力,阶段 II 搭载训练好的 PPO 算法并通过 SAC 训练车辆机组协同能力,降低训练难度与周期。

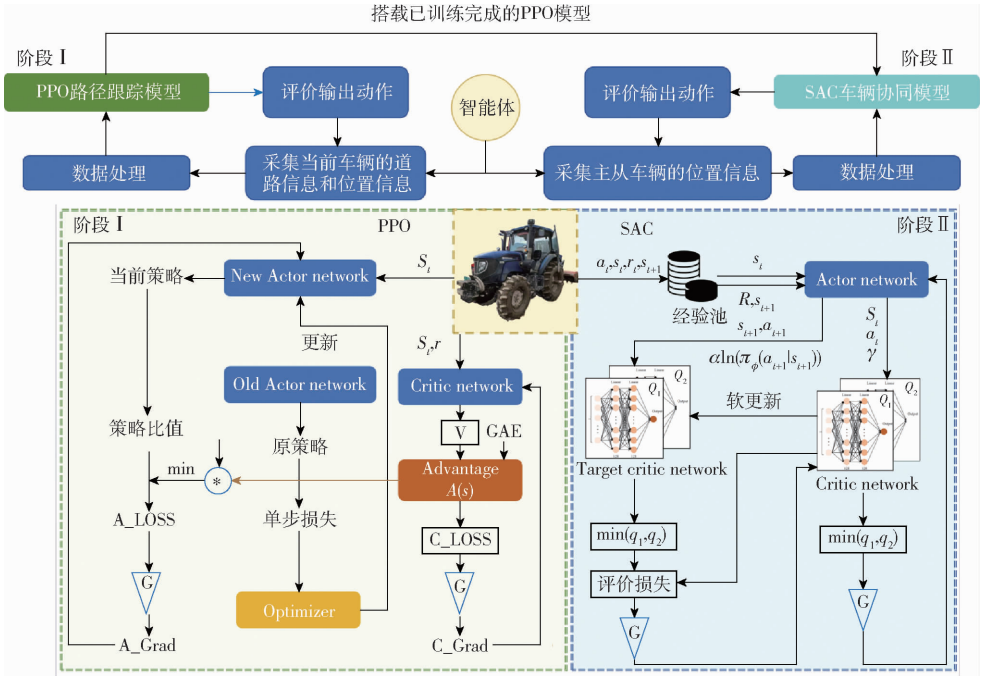


图 3 双环强化学习网络交互框图

Fig. 3 Double loop reinforcement learning network interaction framework

1.4.1.1 融合路径曲率的横向 PPO 算法

当前部分杂交水稻制种田道路弯曲复杂无规则,而授粉作业需要保证机组连续稳定的行驶,要求算法能够快速捕捉并处理位置信息,PPO 作为一种基于策略梯度的算法,能够有效解决传统策略梯度训练不稳定和采样效率低的问题,其目标函数 $L^{CLIP}(\theta_p)$ 为

$$L^{CLIP}(\theta_p) = E_t[\min(r_t(\theta_p), 1 - \epsilon, 1 + \epsilon)A_t] \quad (7)$$

- 式中 E_t ——当前时间步下的数学期望
 θ_p ——策略网络的可学习参数
 $r_t(\theta_p)$ ——新旧策略在同一状态下的动作选择概率比值
 ϵ ——裁剪系数
 A_t ——优势函数

CLIP 操作可限制策略变化范围,防止更新幅度过大。网络包含 2 个 Actor 网络和 1 个 Critic 网络,通过 CLIP 裁剪来限制 New Actor 网络更新幅度。同时,还增加 Advantage 优势函数减少环境随机性导致的误差。将未来路径曲率变化率作为观测状态量之一,使 PPO 提前预判未来路径变化,提升 PPO 网络对弯曲变化路线的适应能力。

1.4.1.2 基于相对位置的纵向 SAC 算法

在完成授粉车辆 PPO 横向路径跟踪训练后,将训练好的转向策略搭载在车辆上进行机组速度协同训练。授粉作业要求授粉机组保持一定位置行驶,确保绳索精准碰击水稻父本。为此,引入 SAC 算法,利用其最大熵框架增强探索能力,适应授粉过程中的动态环境,实现机组协同的速度精准控制,SAC 的总目标函数 $J(\pi)$ 为

$$J(\pi) = \sum_{t=0}^T E_{(s_t, a_t) \sim \rho^\pi} [\gamma^t (r(s_t, a_t) + \alpha H(\pi(\cdot | s_t)))] \quad (8)$$

- 其中 $H(\pi(\cdot | s_t)) = -E[\ln \pi_\phi(a_t | s_t)]$
式中 T ——最大时间步
 t ——时间步索引
 E ——数学期望
 s_t ——第 t 步的环境状态
 a_t ——第 t 步的智能体动作
 ρ^π ——策略 π 下 (s_t, a_t) 的分布
 γ ——折扣因子
 $r(s_t, a_t)$ ——即时奖励函数
 α ——权衡熵与奖励的超参数
 $\pi(\cdot)$ ——智能体策略

$H(\cdot)$ ——熵值

$\pi_\phi(\cdot)$ ——参数化策略网络输出

SAC 采用“软更新”策略,在避免策略更新幅度过大的同时,增强了对速度变化的探索能力。SAC 网络结构包含双 Critic 网络和单 Actor 网络,每次更新选择价值估计较低的 Critic 网络,以提升训练稳定性。将领航车与跟随车的相对位置距离作为观测状态之一,确保其对授粉机组的精确协同控制。

1.4.2 动作空间

为实现更精确的训练与规律特征学习,将转向和速度解耦。横向 PPO 的动作空间设置为转向角 $\delta \in [-\delta_{\max}, \delta_{\max}]$;纵向 SAC 的动作空间设置为速度 $v \in [0, v_{\max}]$ 。训练的网络不断输出 2 个动作的实时计算值实现授粉车辆的协同控制。

1.4.3 观测空间

授粉作业时,依据当前的车辆状态、当前和未来的道路信息可判断下一步该执行的控制动作。因此在双环分层强化学习中,将观测空间表示为

$$s_p = [s_{\text{veh}}, s_{\text{path}}] \quad (9)$$

$$s_s = [s_{\text{veh}}, s_{\text{vr}}, s_{\text{path}}] \quad (10)$$

式中 s_p ——横向路径跟踪所需的状态向量

s_s ——纵向编队协同所需的状态向量

s_{veh} ——领航车或跟随车状态向量

s_{path} ——路径状态向量

s_{vr} ——编队状态向量

在横向 PPO 框架中,其状态空间 s_p 主要由 s_{veh} 和 s_{path} 构成,其中 s_{veh} 包括车辆作业速度 v 、实际航向角 φ ,以及位置坐标 $(x_{\text{veh}}, y_{\text{veh}})$; s_{path} 包括路径横向偏差 l 、路径曲率 κ_0 ,以及未来 N 个点的平均曲率变化率 $\dot{\kappa}_{\text{avg}}$ 。路径横向偏差 l 通过当前车辆位置坐标与参考作业路径的最近距离计算得到,可表示为

$$l = \min_{i=1,2,\dots,n} \left\| \begin{bmatrix} x_{\text{veh}} \\ y_{\text{veh}} \end{bmatrix} - P_i \right\|_2 \quad (11)$$

式中 $\|\cdot\|_2$ ——欧氏距离

P_i ——离散路径的第 i 个路径点

在纵向 SAC 框架中,其状态空间 s_s 主要由 s_{veh} 、 s_{vr} 和 s_{path} 构成,其中 s_{veh} 和 s_{path} 在 PPO 设置基础上增加了 PPO 输出的转向控制 δ ; s_{vr} 引入杂交水稻父本行与机组成员连线的夹角 γ 。

1.4.4 奖励函数

为使强化学习网络能够学习到满足要求的授粉控制动作,需要根据任务特征分别设计相应的奖励函数。

PPO 横向强化学习总奖励 R_p 由导向奖励 R_{dt} 、航向角奖励 R_{ye} 和横向位置误差惩罚 R_{se} 组成,即

$$R_p = R_{dt} + R_{ye} + R_{se} \quad (12)$$

R_{dt} 定义为当前已行驶距离与总路径距离的占比函数,其计算式为

$$R_{dt} = \begin{cases} \frac{s_{\text{cur}}}{s_{\text{tol}}} \lambda_d & \left(\frac{s_{\text{cur}}}{s_{\text{tol}}} < 1 \right) \\ 100 & \left(\frac{s_{\text{cur}}}{s_{\text{tol}}} = 1 \right) \end{cases} \quad (13)$$

式中 s_{cur} ——当前已行驶距离, m

s_{tol} ——路径总长度, m

λ_d ——导向奖励系数

R_{ye} 定义为当前航向角与目标方向夹角的函数,其计算式为

$$R_{ye} = \left(1 - \frac{\gamma_{ye}}{\pi} \right) \lambda_{ye} \quad (14)$$

式中 γ_{ye} ——当前航向角与目标方向的夹角, rad

λ_{ye} ——航向奖励系数

R_{se} 定义为车辆当前位置与路径横向距离的分段函数,其计算式为

$$R_{se} = \begin{cases} \left(1 - \frac{s_{\text{error}}}{0.5} \right) \lambda_{se1} & (s_{\text{error}} \leq 0.2) \\ -\frac{s_{\text{error}}}{0.5} \lambda_{se2} & (0.2 < s_{\text{error}} \leq 0.5) \\ -50 & (s_{\text{error}} > 0.5) \end{cases} \quad (15)$$

式中 λ_{se1} ——误差小于 0.2 m 时的奖励系数

λ_{se2} ——误差在 0.2 ~ 0.5 m 时的奖励系数

s_{error} ——车辆当前的横向距离误差, m

SAC 纵向强化学习网络总奖励 R_s 由导向奖励 R_{dt} 、协同一致性奖励 R_γ 和速度差奖励 R_{vy} 组成,即

$$R_s = R_{dt} + R_\gamma + R_{vy} \quad (16)$$

R_γ 定义为杂交水稻父本行与机组成员连线夹角的函数,其计算式为

$$R_\gamma = \left(1 - \frac{\gamma}{\pi/2} \right) \lambda_\gamma \quad (17)$$

式中 λ_γ ——协同一致性奖励系数

R_{vy} 定义为领航车与跟随车在垂直父本行方向上速度分量比值的函数,其计算式为

$$R_{vy} = \left(1 - \frac{|v_{ly} - v_{fy}|}{\max(v_{ly}, v_{fy})} \right) \lambda_{vy} \quad (18)$$

式中 v_{ly} ——领航车在垂直父本行方向上的速度分量, m/s

v_{fy} ——跟随车在垂直父本行方向上的速度分量, m/s

λ_{vy} ——速度差奖励系数

1.5 模型训练

为了对双环分层强化学习网络进行训练以及验证,采用 Unity 搭建训练场景并进行训练。

1.5.1 训练环境搭建

如图 4 所示,首先从外部导入授粉车辆模型并为其设置碰撞体积,创建一个车辆控制器,实现车辆

速度与转向的底层控制。采用三次样条插值计算进行随机取点拟合生成路径点集,模拟田间的弯曲路线。将 P_0 和 P_3 分别设为起点和终点坐标,位置始终

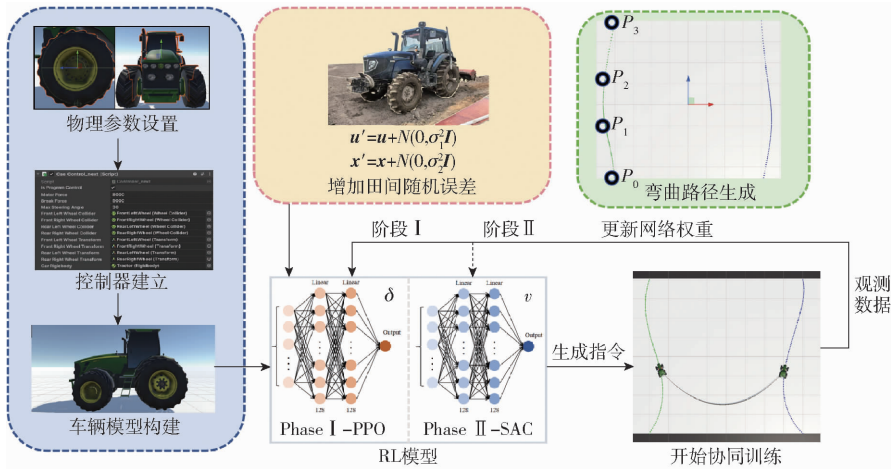


图 4 仿真训练环境结构示意图

Fig.4 Schematic of simulation training environment structure

固定, P_1 和 P_2 设置为随机控制点,该控制点的随机范围 R 设置为

$$R = 0.2 \parallel P_3 - P_0 \parallel \tag{19}$$

则 P_1 和 P_2 的坐标分量取值范围为

$$\begin{cases} x \in [M_x - R, M_x + R] \\ y \in [M_y - R, M_y + R] \\ M = P_0 + 0.5(P_3 - P_0) \end{cases} \tag{20}$$

式中 M ——起点到终点的中点

M_x, M_y ——中点 M 的横、纵坐标

x, y ——随机控制点的横、纵坐标

最终得到车辆的路径点集。通过 Unity 内置的坐标系获取车辆和路径点集的实时坐标,再经过转换后输入到双环分层强化学习网络中。

为使得训练获取的数据更贴近现实数据,在 Unity 中对动作空间变量增加随机高斯噪声

$$\begin{cases} u' = u + N(0, \sigma_1^2 I) \\ x' = x + N(0, \sigma_2^2 I) \end{cases} \tag{21}$$

式中 u' ——添加噪声后的输出控制量

x' ——添加噪声后的观测状态量

u ——输出控制量

x ——观测状态量

σ_1, σ_2 ——噪声标准差

I ——单位矩阵

同理在观测空间中也添加高斯随机噪声,模拟现实传感器的传输波动。噪声波动的获取方法是驾驶已搭载 GNSS 设备和自动驾驶改造的授粉车辆,在非规则田块道路采集。定位误差包括经纬坐标误差和航向角误差,通过静置定点采集 GNSS 模块输出值与地面 GPS 全站仪真值并计算两者的差值得到,执行

误差是采集控制输出转向角与实际执行转向角的偏差、控制输出速度与实际执行速度偏差,采集频率为 10 Hz,每组采集 10 min,采集 3 组,采集完成后计算其标准差增加到仿真环境中作为模拟随机噪声。

在该模拟场景中,每次车辆走完全程或者偏移路径过大时均会随机刷新环境,模拟车辆回归起始点,车辆行驶路径刷新并重新规划,以满足对不同路况路径的适应需求。

1.5.2 网络配置与训练

模型训练基于搭载 Intel Core i9 - 11900K 与 NVIDIA GeForce RTX 4090 的工作站进行。针对授粉车辆路径跟踪与协同控制场景,PPO 与 SAC 均采用 2 层包含 128 神经元的隐藏层;经验缓冲区 PPO 设置为 2 048,SAC 扩大至 100 000 以充分利用离线数据;统一采用线性衰减策略学习率并设置为 0.000 3;同时,为有效限制策略更新幅度并鼓励探索,PPO 裁剪参数设置为 0.2、SAC 熵正则化系数为 0.01。考虑到长期累计奖励的重要性,设置外部奖励信号的折扣因子统一为 0.99。训练最大步数统一设置为 250 000。

路径特征数据基于 1.5.1 节中 Unity 每轮随机生成的弯曲路径,在生成路径的瞬间会获取路径曲率 κ 、平均曲率变化率 $\dot{\kappa}_{avg}$ 等路径数据;干扰误差数据源于实际田间采集的车辆定位误差与控制误差,将这些数据添加到对应仿真环境下的参数中,模拟现实误差;车辆状态数据通过 Unity 实时采集并融合干扰后作为网络输入。网络输出的控制指令设有如下限制:输出速度 $v_{output} \in [0, 7] \text{ m/s}$,输出转向角 $\delta_{output} \in [0, \pi/4] \text{ rad}$,加速度 $|a_{output}| \in [0, 2] \text{ m/s}^2$ 。

1.6 试验设计与验证

1.6.1 仿真对比试验

为体现出双环强化学习网络的跟踪性能和协同性能,将训练好的模型搭载到仿真环境的模拟授粉车辆上,选取 2 组包含 4 个固定点的路径((- 40, - 40)m, (- 20, - 10)m, (- 50, 10)m, (- 40, 40)m)和((40, - 40)m, (15, - 15)m, (45, 20)m, (40, 40)m),通过 Beizer 曲线进行曲线拟合并获取划分好的路径点集,设置领航车的固定速度 $v_l=3.5\text{ m/s}$,最大速度 $v_{\max}=7\text{ m/s}$,转向角 $\delta_{\max}=\pi/4$ 。将双环 PPO-SAC 与单一 PPO、SAC 算法运行的结果进行对比,分析授粉车辆的横向路径跟踪性能与机组协

同效果。

1.6.2 模拟验证试验

为检验双环 PPO-SAC 网络在实际作业情况下的协同能力,进行地面模拟验证试验。试验地点选取在浙江大学紫金港校区(30.300 527°N,120.085 399°E),采用 2 台 Panda-4WD 作为试验车辆,如图 5 所示,通过 GPS 以 10 Hz 的频率进行实时定位并获取相应的航向角、速度等信息;采用 Lora 同样以 10 Hz 的频率进行试验车辆之间的远程无线通信;计算设备包括 2 台 Intel(R) Core(TM) i5-9300H CPU@2.40 GHz 便携式计算机。试验前对各设备进行测试和 PPO-SAC 算法部署。

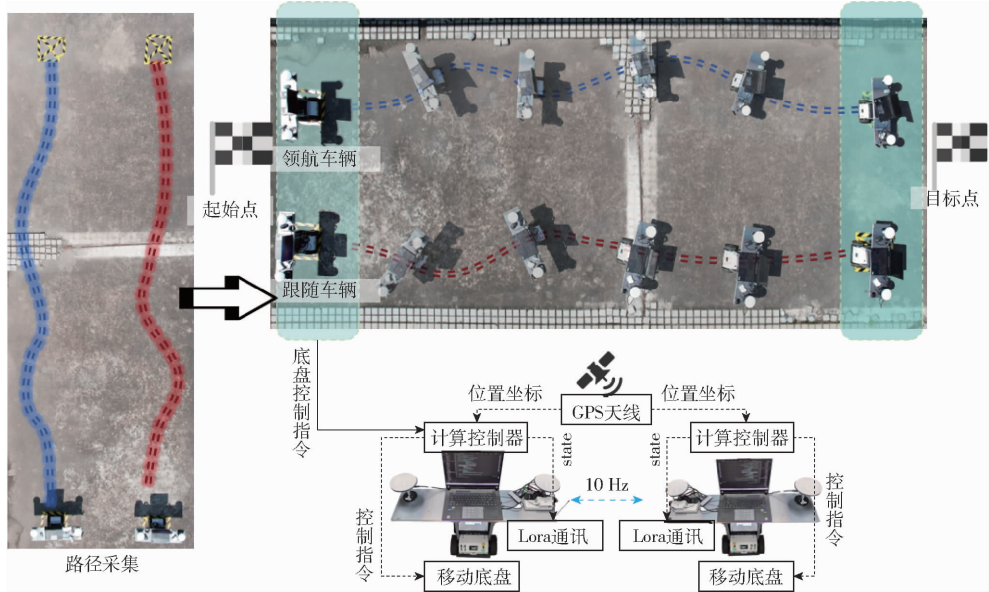


图 5 PPO-SAC 模拟试验示意图
Fig.5 PPO-SAC simulation test schematic

为验证算法的泛化能力,选取 2 条没有任何规律的弯曲轨迹,领航车模拟车辆以速度 0.7 m/s 运行,观察并记录车辆的运行情况。PPO-SAC 作为上层决策控制的核心,仅仅输出转向和速度大小的指令,在地面车辆所采用的设备中已有专门的底层运动控制进行相应指令控制。

2 试验结果与讨论

为评价 PPO-SAC 弯曲路径的机组协同效果,以横向跟踪偏差、航向角偏差为评价路径跟踪性能的指标,以模拟父本行与机组成员连线夹角偏差、机组纵向相对距离偏差作为评价机组协同性能的指标。其中横向跟踪偏差指车辆实际位置到参考路径的垂直距离,航向角偏差指车辆实际航向角与参考路径切线角的差值。基于上述指标定义,将对仿真试验、模拟验证试验及田间试验的测试结果进行详细分析。

2.1 仿真试验结果分析

2.1.1 横向跟踪性能对比

在横向路径跟踪控制中,主要监测了双环 PPO-SAC 算法与单 PPO、SAC 算法的横向跟踪偏差和航向角偏差,如图 6 所示。3 种算法在横向偏差上表现出较大的差异,SAC 误差相比于其他 2 种算法跟踪效果表现较差,跟踪偏差呈现峰值波动变化,最大跟踪偏差达 2.779 m,显然无法满足跟踪作业的需求;而双环 PPO-SAC 算法相比于单 PPO 算法表现得更加平缓稳定,横向行驶平均偏差约为 0.161 m,相比于单一 PPO 降低了 29.9%,航向角偏差平均值为 8.995°,也低于单 PPO 和单 SAC 算法。整体上,在路径跟踪能力上 SAC 表现相对较差,而双环 PPO-SAC 和单 PPO 表现近乎一致,可有效减少跟随车行驶过程中因横向偏移带来的路径偏离风险,同时保持作业过程中与路径的契合度。

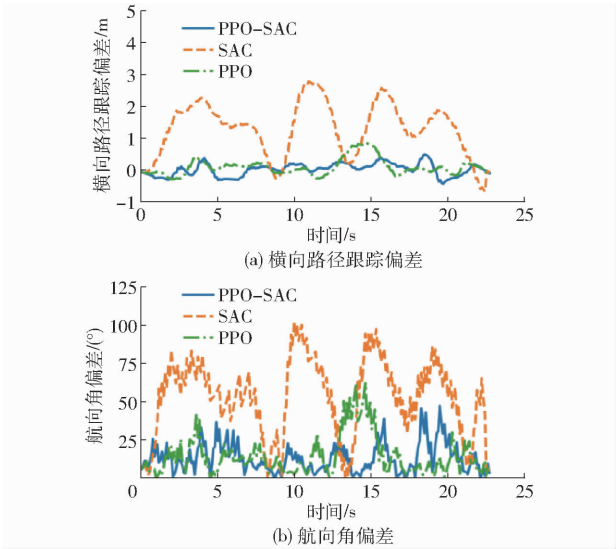


图6 仿真试验横向跟踪曲线

Fig.6 Lateral tracking performance in simulation tests

2.1.2 纵向协同性能对比

在纵向授粉车辆协同控制中(图7a),3种算法的模拟父本行与机组成员连线夹角均保持在 4° 以内,但单一PPO算法的偏差波动幅度大且规律难寻,单一SAC算法波动相对平缓,但仍存在偏差峰值(3.978°),在弯曲较大处协同能力表现较差;双环PPO-SAC算法的平均值为 0.596° 、标准差为 0.417° ,相比于单PPO、单SAC算法误差更小且波动稳定。而在图7b中的机组纵向相对距离中,PPO-SAC距离保持更小,偏差值保持小于 0.739 m ,相比于单SAC算法机组纵向相对距离降低了 22.5% 。总体上双环PPO-SAC算法可以快速感知领航车纵向位置变化并调整跟随车策略,实现更优的偏差控制,提升车辆协同行驶的纵向适配性与整体协调性,协同效

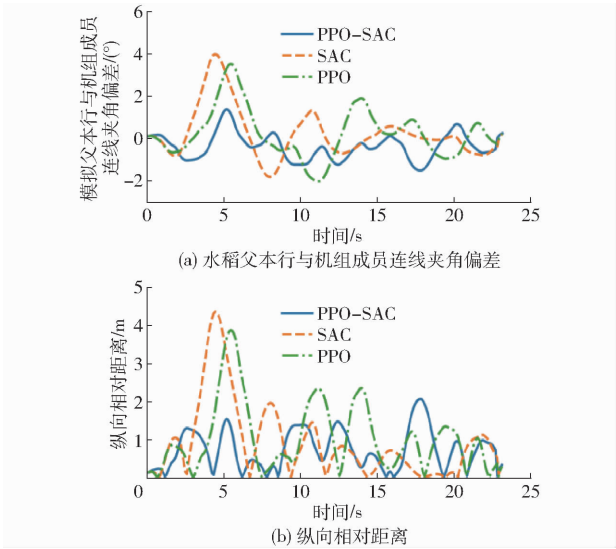


图7 仿真试验纵向协同性能曲线

Fig.7 Longitudinal cooperative performance curves in simulation tests

果优于其余2种单框架的算法。

为了进一步展示双环PPO-SAC调节规律,图8、9分别展示了双环强化学习算法跟随车路径曲率和输出转向角随时间变化曲线以及输出速度与机组纵向相对距离随时间变化曲线,由输出曲线可知,输出转向角曲线与跟随车路径曲率波动高度吻合,在大曲率的区间,如 $2.5\sim7.5\text{ s}$ 和 $12.5\sim17.5\text{ s}$ 时间段中,输出转向角摆动幅度增大;输出速度曲线呈现脉动式变化,即动态的加速和减速,其速度峰值与相对纵向距离峰值呈现近似分布。上述曲线反映出双环强化学习算法在外界环境变化时具有良好的动态响应特性。

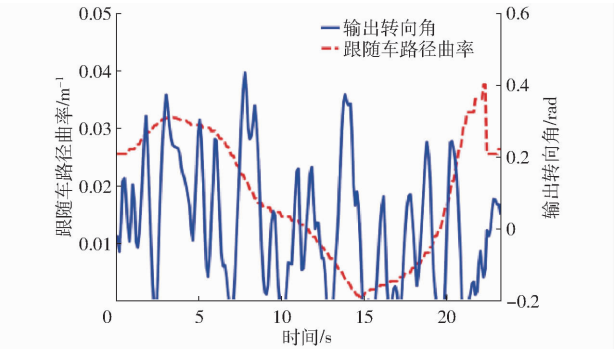


图8 输出转向角与曲率仿真变化图

Fig.8 Simulation curves of output steering angle and path curvature

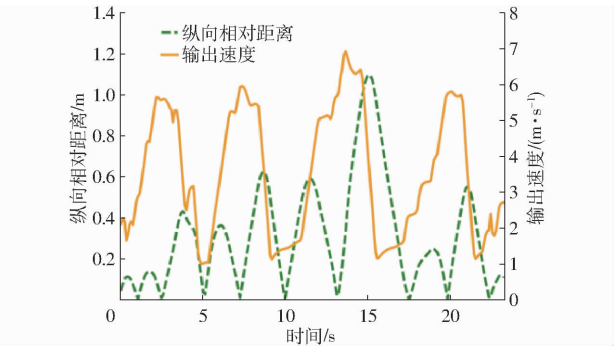


图9 输出速度与机组纵向相对距离仿真变化图

Fig.9 Simulation results of output velocity and longitudinal relative distance between units

通过以上试验,双环PPO-SAC的单一动作针对性训练相比于单一训练算法框架,可以有效减缓对多动作训练难度,在跟踪和协同上表现更为流畅,使得授粉的车辆协同作业在驾驶质量和变速平滑上达到了最佳性能。

2.2 模拟验证试验结果分析

在模拟验证试验中,搭载双环PPO-SAC协同算法的试验车辆顺利完成了机组协同任务,其作业轨迹如图10所示,总计完成时间为 30 s 。

在图11中展示了跟随车在横向上的行驶状态,通过分析可知车辆的横向跟踪偏差控制在 0.393 m

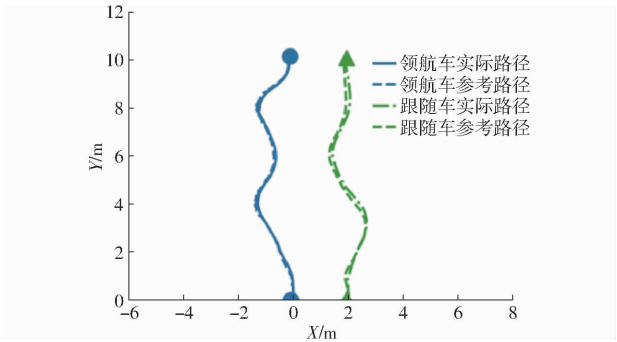


图 10 模拟验证试验车辆协同轨迹

Fig. 10 Vehicle cooperative trajectory in simulation validation

范围内,在转向处会呈现较大波动,但在控制算法干预下会快速进行调整,航向角偏差(向左偏为正;右偏为负)也随路径呈现波动式变换,平均偏差为 9.417° 。

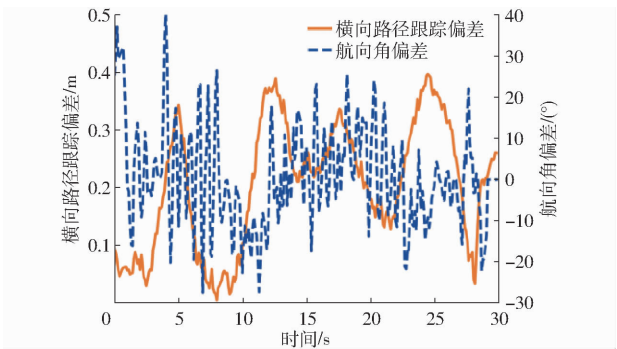


图 11 模拟验证试验横向跟踪性能曲线

Fig. 11 Lateral tracking performance curves in simulation validation

图 12 展示了跟随车在纵向上的行驶状态,在模拟父本行与机组成员连线夹角的变化曲线中,模拟父本行与机组成员连线夹角也呈现波动变化,这表明跟随车的速度在实时变化调节,由于模拟验证试验的两车间距相对较短,因此模拟父本行与机组成员连线夹角相比于仿真环境相对较大,平均值为 11.385° ,但变化规律也更加显著,机组纵向相对距离控制在 $\pm 0.325\text{ m}$ 内,机组纵向平均相对距离为 0.099 m 。

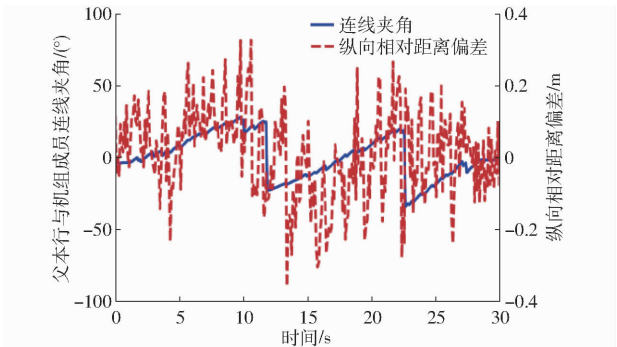


图 12 模拟验证试验纵向协同性能曲线

Fig. 12 Longitudinal coordination performance curves in simulation validation

在图 13、14 中更直观体现出其输出动作的变化情况,通过观测路径曲率和曲率变化,自主调节转向

角进行实时路径跟踪,图中输出转向角变化趋势与路径曲率基本一致,由于 GPS 定位点的漂移和变化,此处的曲率呈现波动变化。在跟随车输出速度与机组纵向相对距离变化图中,当相对位置增大时(如 $5\sim 10\text{ s}$ 处),跟随车即增大输出速度(如 $7.5\sim 12.5\text{ s}$ 处),以最大输出速度 1.36 m/s 对领航车进行追赶,保证协同一致性,机组最大纵向相对距离为 0.325 m ,协同一致性良好。

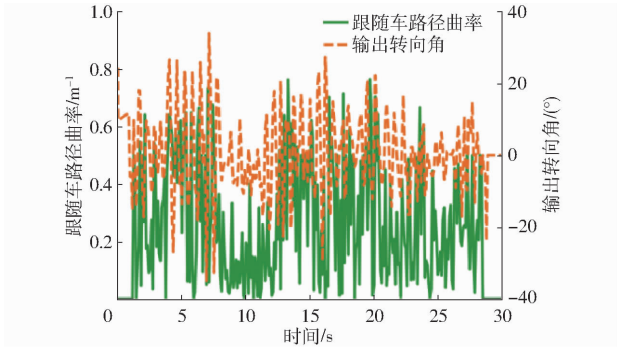


图 13 模拟验证试验输出转向与曲率变化曲线

Fig. 13 Variation of output steering angle and curvature in simulation verification tests

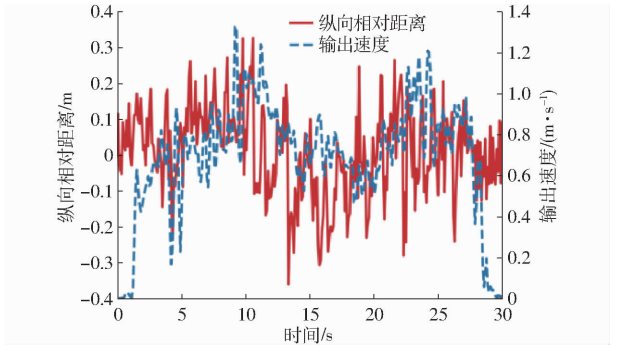


图 14 模拟验证试验输出速度与相对位置变化曲线

Fig. 14 Variation of output velocity and relative position in simulation verification tests

2.3 田间试验验证

为了验证 PPO-SAC 算法在真实杂交水稻制种授粉时的作业效果,进行了杂交水稻制种机组田间试验验证。试验装备包括 2 台经过自动驾驶改造的阿克曼底盘水稻插秧机,包含电控方向盘、电子油门以及控制器,2 台车辆之间通过 Lora 进行远程无线通讯。试验地点选在上海市青浦区联适生态智慧农场青浦基地,选取的试验区域是一片符合农田作业条件的田块。图 15 展示了领航车与跟随车的作业轨迹,通过监测跟随车的作业状态来判断该算法在弯曲路径的协同性能。

图 16 和图 17 分别展示了跟随车对目标路径的跟踪偏差以及与领航车辆的机组协同表现。由图 16 可知,搭载 PPO-SAC 算法的跟随车辆最大跟踪偏差小于 0.416 m ,平均跟踪偏差为 0.139 m ,航

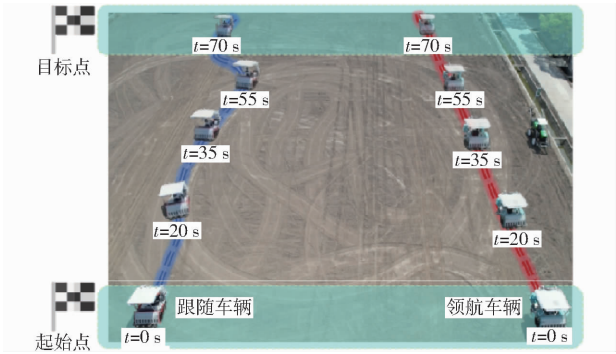


图 15 田间试验作业示意图

Fig. 15 Schematic of field test operation

向角偏差基本能控制在 $\pm 30^\circ$ 以内,跟踪效果稳定;在纵向协同偏差表现方面,由图 17 可知,模拟父本行与机组成员连线夹角与机纵纵向相对距离偏差曲线基本吻合,在 55~65 s 处出现较大波动,对应位置恰好是道路弯曲较大处,而纵向偏差最大达 1.470 m,这是因为系统设置的最低作业速度限制,跟随车在转向时,领航车仍然以最低速度行驶,但是会在跟随恢复直行时快速调整,其机组平均纵向相对距离偏差为 0.585 m,协同效果良好,能够满足杂交水稻制种授粉作业的需求。

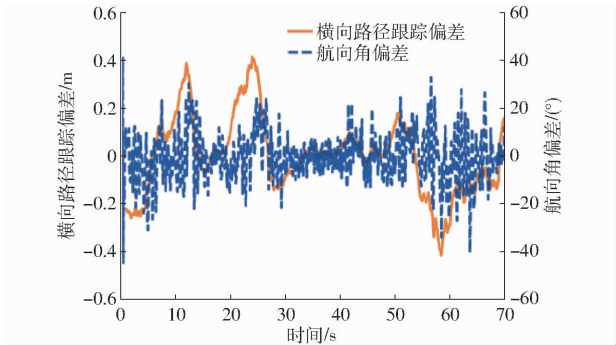


图 16 田间试验横向跟踪性能曲线

Fig. 16 Lateral tracking performance in field tests

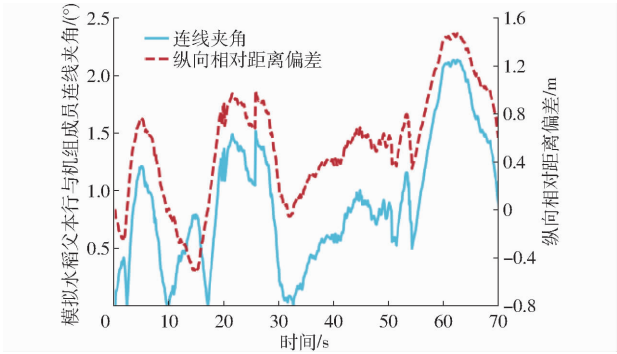


图 17 田间试验纵向协同性能曲线

Fig. 17 Longitudinal cooperative performance in field tests

3 结论

(1) 提出了纵横向解耦双环强化学习 PPO-SAC 框架,将机组协同授粉拆分为路径跟踪与协同控制 2 个子任务,并分别融合路径曲率与机纵纵向相对距离,解决了单一算法处理多任务时的训练困难、执行精度低的问题,实现对路径的动态自适应跟踪和实时协同。

(2) 融合路径曲率信息的横向路径跟踪的 PPO 算法,平均跟踪偏差为 0.161 m,相比于单 PPO 框架降低了 29.9%;融合相对位置信息与 PPO 输出转向角的纵向协同控制 SAC 框架纵向平均相对偏差为 0.739 m,相比于单 SAC 算法减少了 22.5%。

(3) 集成双环 PPO-SAC 算法的地面机组,路径的跟踪偏差小于 0.393 m,机纵纵向相对距离偏差小于 0.325 m,跟随车输出的转向和速度表明该算法可以根据路径曲率和领航车位置进行有效的预测;田间试验中机组平均协同偏差为 0.585 m,能够较好满足田间授粉作业的需求。

参 考 文 献

[1] Feng H, Xi T, Ruan Y, et al. Topological organization for hybrid rice growth stages phenotype based on contrastive clustering [J]. Expert Systems with Applications, 2025, 264: 125859.

[2] Zeng D, Tian Z, Rao Y, et al. Rational design of high-yield and superior-quality rice [J]. Nat Plants, 2017, 3(4): 17031.

[3] 潘宇镭,吴雨铨,李成龙,等. 杂交水稻制种父本倒播差插秧视觉导航线实时提取方法[J]. 农业机械学报, 2024, 55(增刊 1): 41-50.

Pan Yulei, Wu Yuhua, Li Chenglong, et al. Real-time extraction method of visual navigation line for reverse sowing difference transplanting of male parent in hybrid rice seed production [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(S1): 41-50. (in Chinese)

[4] 姜锐,林键沁,林宗辉,等. 航时可延展水稻辅助授粉无人机设计与作业参数优选[J]. 农业机械学报, 2025, 56(2): 229-239.

Jiang Rui, Lin Jianqin, Lin Zonghui, et al. Design of rice supplementary pollination unmanned aerial vehicle with extendable endurance and optimization of operation parameters [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(2): 229-239. (in Chinese)

[5] Jiang Q, Wang Y, Chen J, et al. Optimizing the working performance of a pollination machine for hybrid rice [J]. Computers and Electronics in Agriculture, 2021, 187(4): 106282.

[6] Chamer A M, Medan D, Mantese A I, et al. Impact of pollination on sunflower yield: is pollen amount or pollen quality what

- matters? [J]. *Field Crops Research*, 2015, 176: 61–70.
- [7] 王永维,何焯亮,陈军,等. 碰撞气吹式杂交水稻授粉机结构与参数优化[J]. *浙江大学学报(农业与生命科学版)*, 2018, 44(1): 98–106.
- Wang Yongwei, He Zhuoliang, Chen Jun, et al. Optimization on structure and parameters of a collision-pneumatic hybrid rice pollination machine[J]. *Journal of Zhejiang University (Agric. & Life Sci.)*, 2018, 44(1): 98–106. (in Chinese)
- [8] Xi T, Feng H, Wang Y. Optimizing the working parameters of pneumatic pollinators based on the kinematic properties of rice panicles[J]. *Journal of the ASABE*, 2023, 66(1): 65–74.
- [9] Xi T, Sun L, Wang Y, et al. Optimizing the airflow velocity combinations acting on male parent rows for hybrid rice pollination [J]. *Journal of the ASABE*, 2023, 66(1): 13–22.
- [10] 翁晓星,徐锦大,赵晋,等. 多旋翼无人机辅助籼粳杂交稻制种授粉研究[J]. *中国农机化学报*, 2022, 43(9): 24–29.
- Weng Xiaoxing, Xu Jinda, Zhao Jin, et al. Research on multi-rotor UAV assisted pollination for indica-japonica hybrid rice seed production[J]. *Journal of Chinese Agricultural Mechanization*, 2022, 43(9): 24–29. (in Chinese)
- [11] 李中秋,汤楚宙,王慧敏,等. 气流作用位置对杂交水稻制种气力式授粉花粉分布的影响[J]. *农业工程学报*, 2012, 28(增刊2): 107–113.
- Li Zhongqiu, Tang Chuzhou, Wang Huimin, et al. Effects of location of airflow on pollen distribution for pneumatic pollination in hybrid rice breeding[J]. *Transactions of the CSAE*, 2012, 28(S2): 107–113. (in Chinese)
- [12] Xi T, Shi R, Feng H, et al. TD-CFD-DPM coupled method for multi-objective optimization of collision pollination parameters in hybrid rice seed production[J]. *Smart Agricultural Technology*, 2025, 12: 101130.
- [13] 沈跃,张亚飞,刘慧,等. 农业装备自动控制技术研究综述[J]. *农业机械学报*, 2023, 54(8): 1–18.
- Shen Yue, Zhang Yafei, Liu Hui, et al. Research review of agricultural equipment automatic control technology [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(8): 1–18. (in Chinese)
- [14] Xie B, Jin Y, Faheem M, et al. Research progress of autonomous navigation technology for multi-agricultural scenes [J]. *Computers and Electronics in Agriculture*, 2023, 211: 107963.
- [15] Yu F, Xu T, Guo Z, et al. Research status and prospects of key technologies for rice smart unmanned farms [J]. *Smart Agriculture*, 2024, 6(6): 1–22.
- [16] 金诚谦,陈钧龙,刘政,等. 收获卸粮协同作业位速耦合控制方法[J]. *农业机械学报*, 2025, 56(9): 427–439.
- Jin Chengqian, Chen Junlong, Liu Zheng, et al. Position-velocity coupling control method for coordinated operation of harvesting and grain unloading [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2025, 56(9): 427–439. (in Chinese)
- [17] 窦汉杰,陈震宇,翟长远,等. 果园智能化作业装备自主导航技术研究进展[J]. *农业机械学报*, 2024, 55(4): 1–22.
- Dou Hanjie, Chen Zhengyu, Zhai Changyuan, et al. Research progress on autonomous navigation technology of intelligent orchard operation equipment [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(4): 1–22. (in Chinese)
- [18] 郭万金,李儒,郝钦磊,等. 农业物料移运机器人协同作业时间最优轨迹规划方法[J]. *农业机械学报*, 2024, 55(1): 22–38.
- Guo Wanjin, Li Ru, Hao Qinlei, et al. Time-optimal trajectory planning method for coordinated operation of agricultural material handling robots [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(1): 22–38. (in Chinese)
- [19] Ding F, Zhang W, Luo X, et al. Gain self-adjusting single neuron PID control method and experiments for longitudinal relative position of harvester and transport vehicle [J]. *Computers and Electronics in Agriculture*, 2023, 213: 108215.
- [20] Zhang C, Noguchi N. Development of a multi-robot tractor system for agriculture field work [J]. *Computers and Electronics in Agriculture*, 2017, 142(Part A): 79–90.
- [21] Zhang C, Noguchi N, Yang L. Leader-follower system using two robot tractors to improve work efficiency [J]. *Computers and Electronics in Agriculture*, 2016, 121: 269–281.
- [22] 魏新华,邓屹,崔鑫宇,等. 基于模糊算法的农机编队转场多机协同控制方法[J]. *农业机械学报*, 2025, 56(2): 48–60.
- Wei Xinhua, Deng Yi, Cui Xinyu, et al. Multi-machine coordinated control method for agricultural machinery formation transition based on fuzzy algorithm [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2025, 56(2): 48–60. (in Chinese)
- [23] Cheng J, Zhang B, Zhang C, et al. A model-free adaptive predictive path-tracking controller with PID terms for tractors [J]. *Biosystems Engineering*, 2024, 242: 38–49.
- [24] Liu J, Wu X, Quan L, et al. Fuzzy adaptive PID control for path tracking of field intelligent weeding machine [J]. *Aip Advances*, 2024, 14: 035045.
- [25] Liu L, Wang X, Xie J, et al. Path planning and tracking control of orchard wheel mower based on BL-ACO and GO-SMC [J]. *Computers and Electronics in Agriculture*, 2025, 228: 109696.
- [26] Liu L, Wang X, Wang X, et al. Path planning and tracking control of tracked agricultural machinery based on improved A* and fuzzy control [J]. *Electronics*, 2024, 13(1): 188–209.

Yang Tong, Han Jianyou. Rigid-body guidance mechanism synthesis through four mixed positions [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(3): 203–207, 196. (in Chinese)

[16] 曹阳, 韩建友. 基于解域的四位置空间 RCCC 机构综合[J]. 农业机械学报, 2016, 47(8): 399–405.
Cao Yang, Han Jianyou. Synthesis of RCCC linkage to visit four given positions based on solution region [J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(8): 399–405. (in Chinese)

[17] 钱卫香, 韩建友. 四位置导引机构解域分析及稳健性评价方法[J]. 机械设计与研究, 2008, 24(6): 17–21.
Qian Weixiang, Han Jianyou. Solution region analysis and robust penalty method of 4-position rigid body guidance [J]. Mechanical Design and Research, 2008, 24(6): 17–21. (in Chinese)

[18] Dhawan S, Kapoor S, Kumar S. Numerical method for advection diffusion equation using FEM and B-splines [J]. Journal of Computational Science, 2012, 3(5): 429–437.

[19] Arora G, Singh B K. Numerical solution of Burgers' equation with modified cubic B-spline differential quadrature method [J]. Applied Mathematics and Computation, 2013, 224: 166–177.

[20] 李袁迪, 张立强, 郭鹏程, 等. 基于 ADAMS 的油茶嫁接机关键结构设计及仿真[J]. 农机化研究, 2025, 47(7): 205–212.
Li Yuandi, Zhang Liqiang, Guo Pengcheng, et al. Design and simulation of key structures of a *Camellia oleifera* grafting machine based on ADAMS [J]. Journal of Agricultural Mechanization Research, 2025, 47(7): 205–212. (in Chinese)

[21] 高志远, 晏芙蓉. 基于 ANSYS Workbench 的行星齿轮减速器仿真分析[J]. 机电工程技术, 2024, 53(9): 104–107.
Gao Zhiyuan, Yan Furong. Simulation analysis of a planetary gear reducer based on ANSYS Workbench [J]. Mechanical & Electrical Engineering Technology, 2024, 53(9): 104–107. (in Chinese)

[22] 徐建勇, 严辉, 周永光. 缝纫机针杆刺布机构的惯性力优化平衡研究[J]. 科技创新与应用, 2021, 11(30): 1–6.

[23] 孙杨波, 高媛, 钱峰, 等. 高速压力机机构综合平衡优化设计[J]. 机械设计与制造, 2021(7): 1–4.
Sun Yangbo, Gao Yuan, Qian Feng, et al. Optimization design on comprehensive balance of high-speed press mechanism [J]. Mechanical Design and Manufacturing, 2021(7): 1–4. (in Chinese)

[24] 刘家成, 葛正浩, 李杰, 等. 二级往复小型空压机动平衡优化设计与试验验证[J]. 机械强度, 2023, 45(6): 1445–1451.
Liu Jiacheng, Ge Zhenghao, Li Jie, et al. Optimal design of dynamic balance of two-stage reciprocating small air compressor and coupling vibration test analysis of motor [J]. Mechanical Strength, 2023, 45(6): 1445–1451. (in Chinese)

~~~~~

(上接第 225 页)

[27] 黄文岳, 魏新华, 汪岸哲, 等. 基于模糊快速滑模的农用牵引车辆直线路径跟踪控制[J]. 农业机械学报, 2024, 55(2): 415–422.  
Huang Wen Yue, Wei Xin Hua, Wang An Zhe, et al. Straight-line path tracking control of agricultural traction vehicles based on fuzzy fast sliding mode [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 415–422. (in Chinese)

[28] Han M, He H, Cao J, et al. Learning based model predictive path tracking control for autonomous buses [C] // Intelligent Vehicles Symposium (IV), IEEE, 2024: 2680–2687.

[29] Gao W, Mynuddin M, Wunsch D C, et al. Reinforcement learning-based cooperative optimal output regulation via distributed adaptive internal model [J]. IEEE transactions on Neural Networks and Learning Systems, 2022, 33(10): 5229–5240.

[30] 祝青园, 程家琪, 陈轩伟, 等. 无人驾驶铰接转向车辆路径跟踪控制研究综述[J]. 农业机械学报, 2024, 55(1): 1–21.  
Zhu Qingyuan, Cheng Jiaqi, Chen Xuanwei, et al. A review of path tracking control for unmanned articulated steering vehicles [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 1–21. (in Chinese)

[31] Li M, Cao Z, Li Z. A reinforcement learning-based vehicle platoon control strategy for reducing energy consumption in traffic oscillations [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(12): 5309–5322.

[32] Wang P, Li H, Chan C Y. Continuous control for automated lane change behavior based on deep deterministic policy gradient algorithm [C] // Intelligent Vehicles Symposium (IV), IEEE, 2019: 1454–1460.

[33] Liu Y, Liu Z, Wang G, et al. Flexible multi-UAV formation control via integrating deep reinforcement learning and affine transformations [J]. Aerospace Science and Technology, 2025, 157: 109812.

[34] Obradovic J, Krizmancic M, Bogdan S. Decentralized multi-robot formation control using reinforcement learning [C] // Sarajevo, Bosnia and Herzegovina, IEEE, 2023: 1–7.

[35] Wang H, Li J, Tao H, et al. Autonomous dynamic formation for maritime target tracking using multi-agent reinforcement learning [J]. Engineering Applications of Artificial Intelligence, 2025, 154: 110904.