DOI:10.19651/j.cnki.emt.2518291

YOLO 检测下基于 ETC-DDPG 算法的无人机视觉跟踪*

庄晶颖[1] 刘 磊[1] 闫冬梅[2] 梁成庆[3]
(1.河海大学数学学院 南京 211100; 2.南京邮电大学现代邮政学院 南京 210003;
3.河海大学人工智能与自动化学院 常州 213000)

摘 要: 为提升无人机动态目标跟踪效率和精度,提出结合 DDPG 算法和 YOLO 目标检测技术的 ETC-DDPG 算法。该算法引入事件触发机制,通过动态调整策略更新频率来提高算法的决策效率;引入课程学习构建分阶段训练框架,逐步提升无人机对复杂任务的适应性。实验结果表明,ETC-DDPG 算法能够有效提升动态目标跟踪任务的跟踪效率和训练过程稳定性,并能减少计算资源需求,成功率可达 93.357%,相比原始 DDPG 算法和 ETC-TD3 算法各项指标都有所提升,其中成功率分别提升 56.175%和 37.1%,并通过消融实验验证了事件触发机制和课程学习的协同作用,为无人机的自主执行任务提供了参考。

关键词: 无人机;事件触发机制;课程学习;视觉跟踪;强化学习
中图分类号: TP273; TP18; TP391.9; TN957 **文献标识码**: A **国家标准学科分类代码**: 520.60

Event-triggered curriculum DDPG for UAV visual tracking with YOLO detection

Zhuang Jingying[1] Liu Lei[1] Yan Dongmei[2] Liang Chengqing[3]
(1. School of Mathematics, Hohai University, Nanjing 211100, China;
2. School of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
3. College of Artificial Intelligence and Automation, Hohai University, Changzhou 213000, China)

Abstract: This paper proposed an Event-triggered Curriculum DDPG algorithm to improve the efficiency and accuracy of dynamic target tracking for UAVs. The algorithm combined Deep Deterministic Policy Gradient (DDPG) and YOLO object detection technology. It introduced an event-triggered mechanism to dynamically adjust the policy update frequency, enhancing decision-making efficiency. Additionally, it incorporated curriculum learning to create a staged training framework, gradually improving the UAV's adaptability to complex tasks. Experimental results showed that the ETC-DDPG algorithm effectively improved the tracking efficiency of dynamic target tracking task and the stability of training process, and reduced the demand for computing resources, achieving a success rate of 93.357%. Compared with the original-DDPG algorithm and ETC-TD3 algorithm, the success rate is improved by 56.175% and 37.1% respectively. The collaborative effect of the event-triggered mechanism and curriculum learning was verified by ablation experiment, providing a reference for autonomous task execution in UAVs.

Keywords: UAV; event-triggered mechanism; curriculum learning; visual tracking; reinforcement learning

0 引 言

随着科技发展,无人机(unmanned aerial vehicle, UAV)因其机动性强、成本低等优势,被广泛应用于搜索巡逻[1]、电力巡检[2]等任务[3-4]。其中,对地面车辆的稳定跟踪是无人机执行自主任务的关键能力[4]。比如,在搜索巡逻中,无人机需要实时监控目标车辆的位置和动态;电力巡检中,跟踪地面车辆有助于提高效率和覆盖范围。因此,推动无人机跟踪车辆在各类实际应用中的部署至关重要。

目前,针对无人机跟踪车辆的主要方法[5-6]包括视觉跟踪、雷达跟踪以及融合定位导航技术等。相比雷达在高精度目标定位和辨识方面的局限、融合定位导航的复杂实时

收稿日期:2025-03-09
*基金项目:航空科学基金(2024Z071108001)、教育部重点实验室开放基金(Scip20240111)、安徽省普通高校交通信息与安全重点实验室开放课题(KLAHEI180188)、中央高校业务费(B240203012)项目资助

计算对系统实时性和响应速度的影响，视觉方法依托于轻量、低成本的视觉传感器，能够高精度地捕捉目标外观特征并获取丰富的环境信息，在目标识别与实时性方面具有显著优势。基于视觉方法，文献[7]提出轻量级多车辆跟踪模型，结合行为信息和视觉信息，通过轨迹预测等模型，实现实时车辆跟踪。文献[8]提出基于交并比匹配的多目标跟踪方法，通过结合传统目标检测技术和运动信息，提高了无人机视频跟踪车辆的性能。基于视觉的无人机跟踪技术在车辆跟踪和目标识别等方面的优势，相关研究正在不断深化[9-10]。为进一步提升视觉跟踪技术的性能，近年来相关研究开始将深度强化学习与视觉图像相结合[11]，以增强无人机在动态环境中的自适应决策能力。

深度强化学习通过与环境交互优化策略，能够显著提升无人机在动态环境中的自适应决策能力，结合视觉图像后，能够进一步实现无人机的精准跟踪与协同。文献[12]提出基于视觉感知和深度强化学习的无人机端到端主动目标跟踪控制方法，采用任务分解和预训练的迁移学习策略，达到稳定跟踪车辆等移动目标的效果。文献[13]提出结合深度学习和强化学习的框架，利用着陆视觉系统和内存整合双延迟深度确定性策略梯度算法(twin delayed deep deterministic policy gradient，TD3)实现四旋翼无人机跟踪无人车。文献[14]则通过结合 YOLO 视觉检测与深度确定性策略梯度算法(deep deterministic policy gradient，DDPG)控制策略，将控制策略部署为级联位置控制器，从而在硬件与软件上成功实现无人机对地面车辆的自主跟踪。

尽管结合视觉图像与深度强化学习在无人机自主跟踪领域取得了显著进展，但仍存在训练过程不稳定、计算资源需求高等局限。课程学习[15]通过分阶段训练策略，从简单任务逐步过渡到复杂任务，能够有效解决深度强化学习训练过程不稳定的问题。文献[16]和[17]在无人机强化学习训练任务中结合课程学习分阶段训练策略，解决了模型训练速度慢、计算量大和响应不及时的问题。由于图像数据的获取和处理需要时间，结合视觉的深度强化学习训练过程较慢。事件触发机制通过动态调整策略更新频率，仅在环境状态显著变化时触发学习过程，避免了无效计算，显著降低了计算资源需求并提高了算法效率[18]。文献[19]提出了一种基于事件驱动控制的多智能体编队强化学习算法，通过事件触发条件动态更新动作决策，保持系统性能的同时有效降低了计算和通信资源消耗。

基于上述研究，本文针对无人机自主跟踪系统计算资源需求高、训练过程不稳定的问题，提出了一种基于事件触发机制和课程学习的 ETC-DDPG 算法(event-triggered curriculum DDPG)，通过动态调整策略更新频率和分阶段训练策略，降低了计算资源需求并提高了训练稳定性。实验结果表明，ETC-DDPG 算法显著提升了无人机在动态环境中的跟踪性能，不仅提升了无人机的跟踪精度，还优化了算法计算效率和响应速度，表现出更好的鲁棒性和稳定性。

1　跟踪算法和算法优化

本文旨在通过 ETC-DDPG 算法训练四旋翼无人机，使其能够自主跟随一辆运动轨迹未知的地面无人车。无人机根据 YOLO 检测提供的视觉反馈，实时获取地面无人车的位置信息，调整飞行轨迹，确保地面无人车始终保持在视野中心，跟踪任务示意图如图 1 所示。本节具体说明跟踪任务所使用的算法和优化方法。

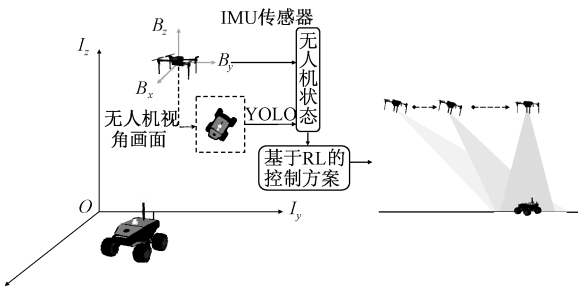


图 1　无人机跟踪任务示意图

Fig. 1　Schematic diagram of drone tracking task

1.1　基于 YOLO 的目标检测

YOLO[20]是一种将目标检测转化为回归问题的深度学习模型。在无人机跟踪任务中，目标通常处于快速运动状态，YOLO 的高检测速度能够确保无人机实时捕捉目标位置。因此本文采用 YOLO 对车辆进行检测。

基于 YOLO 的目标检测核心思想如下：输入图像 $\boldsymbol{I} \in \mathbb{R}^{H\times W\times 3}$，通过卷积神经网络处理后输出一个大小为 $S\times S\times(B\times 5+C)$ 的张量。其中，H、W 和 3 分别是图像高度和宽度和颜色通道数(RGB)，$S\times S$ 表示将图像划分为 $S\times S$ 个网格，B 是每个网格预测的边界框数量，C 是目标的类别数量，5 是每个边界框的预测值，包括边界框的中心坐标 (x,y)、宽度和高度 (w,h) 及置信度 $p=P\times IoU_{pred}$，P 是该网格包含某个物体的概率，IoU_{pred} 是预测框和真实框之间的交并比(intersection over union，IoU)。

每个网格还会预测 C 个类别的概率分布 $P(c\mid x)$，表示该网格内检测到不同类别物体的条件概率。假设网格内包含某个目标物体，则类别的条件概率为：

$$P(c\mid x)=\frac{e^{score_c}}{\sum_{i=1}^{C}e^{score_i}} \tag{1}$$

其中，$score_c$ 是该网格内预测物体属于类别 c 的得分。

YOLO 的训练目标是最小化以下损失函数：

$$\mathcal{L}=\sum_{i=1}^{S^2}\sum_{j=1}^{B}1_i^{obj}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2+(w_i-\hat{w}_i)^2+(h_i-\hat{h}_i)^2+(p_i-\hat{p}_i)^2]+\lambda_{noobj}\sum_{i=1}^{S^2}\sum_{j=1}^{B}1_i^{obj}(p_i-\hat{p}_i)^2 \tag{2}$$

其中，1_i^{obj} 是指示函数，若第 i 个网格包含物体则为 1，否则为 0。$\hat{x}_i$、$\hat{y}_i$、$\hat{w}_i$、$\hat{h}_i$、$\hat{p}_i$ 是目标边界框的真实值。λ_{noobj} 是

无物体的损失加权系数,避免对无目标区域进行过多训练。

1.2 DDPG 算法

无人机跟踪任务需要对无人机的连续控制和高精度决策。DDPG 算法[21]结合确定性策略梯度和 Actor-Critic 框架,专为处理连续动作空间的任务而设计,在无人机跟踪任务中具有显著优势。

在 DDPG 算法中,Actor 网络的任务是学习策略函数 $\mu(s_t \mid \theta_\mu)$,在状态 s_t 下选择最佳动作 a_t,根据 Critic 的 Q 值反馈,通过梯度下降法更新优化自身网络,更新公式如式(3)所示。

$$\nabla_{\theta_\mu} J(\theta_\mu) = \mathbb{E}[\nabla_a Q(s_t, a_t \mid \theta_Q) \nabla_{\theta_\mu} \mu(s_t \mid \theta_\mu)] = \frac{1}{N} \sum_{i=1}^{N} \nabla_{\theta_\mu} \mu(s_i \mid \theta_\mu) \nabla_a Q_{\theta_Q}(s_i, a_i) \mid_{a_i = \mu(s_i \mid \theta_\mu)} \tag{3}$$

Critic 网络负责估算给定状态-动作对 (s_t, a_t) 的预期总回报 $Q(s_t, a_t \mid \theta_Q)$,并通过最小化损失函数 $L(\theta_Q)$ 来更新参数,损失函数如式(4)所示。

$$L(\theta_Q) = \mathbb{E}[(Q(s_t, a_t \mid \theta_Q) - y)^2] = \frac{1}{N} \sum_i (r_i + \gamma Q_{\theta_Q}(s_{t+1}, \mu(s_{t+1})) - Q_{\theta_Q}(s_i, a_i))^2 \tag{4}$$

其中,N 为批采样大小,$J(\theta_\mu)$ 是期望回报,θ_μ 和 θ_Q 分别为 Actor 和 Critic 网络参数,$y = r_t + \gamma Q_{\theta_Q}(s_{t+1}, \mu(s_{t+1}))$ 为目标 Q 值。

为了增强训练的稳定性,DDPG 引入目标网络用于计算目标 Q 值和更新策略,其参数 θ'_μ 和 θ'_Q 会通过软更新的方式逐步逼近主网络的参数 θ_μ 和 θ_Q,避免训练过程中的不稳定性。目标网络的参数更新公式如式(5)所示。

$$\begin{cases} \theta'_\mu \leftarrow \tau\theta_\mu + (1-\tau)\theta'_\mu \\ \theta'_Q \leftarrow \tau\theta_Q + (1-\tau)\theta'_Q \end{cases} \tag{5}$$

其中,τ 是目标网络更新系数。

由于 DDPG 需要同时训练 Actor 网络和 Critic 网络,且经验回放机制需要存储大量历史数据,计算资源需求较高,而无人机搭载的计算资源有限,故本文引入事件触发机制和课程学习,进一步提升 DDPG 在跟踪任务中的性能。

1.3 基于事件触发机制的算法优化

事件触发机制是一种控制策略,在这种机制下,当状态变量超出预定阈值时,控制更新才会被触发。将事件触发机制应用于强化学习的环境控制和决策更新中可以避免频繁的无效更新,减少不必要的计算和交互,降低探索性动作的成本。在本文中,只有当目标无人车的状态 s 发生较大变化时,才会触发学习更新。系统无需对每一帧图像进行计算和学习更新,从而大幅提高了学习效率并减少了视觉数据处理和计算的负担。

将无人机从 $t-1 \sim t$ 时刻的状态偏差值记为:

$$e(t) = \frac{\mid S_t(d,a,v) - S_{t-1}(d,a,v) \mid}{\mid S_{t-1}(d,a,v) \mid} = \sqrt{\omega_d \left(\frac{d_t - d_{t-1}}{d_{t-1}}\right)^2 + \omega_a \left(\frac{a_t - a_{t-1}}{a_{t-1}}\right)^2 + \omega_v \left(\frac{v_t - v_{t-1}}{v_{t-1}}\right)^2} \tag{6}$$

其中,$t > 0$ 为时刻,$S_t(d,a,v)$ 为 t 时刻的状态,$S_{t-1}(d,a,v)$ 为 $t-1$ 时刻的状态,d 为与目标无人机的距离,a 为无人机加速度,v 为速度,ω 为各项权重。当 $e(t) > \varepsilon$ 时,触发后续的动作更新。新动作是基于策略模型计算出来的,并加入探索噪声,最后对动作进行裁剪,确保动作在有效范围内。在本文中,考虑到无人机本身计算资源受限,为减小计算资源消耗,故采用基于任务需求的静态阈值 $\varepsilon = 0.1$。

1.4 基于课程学习的算法优化

强化学习中,智能体在训练过程中常常面临陷入局部最优解或训练不稳定等挑战。课程强化学习将复杂任务逐步拆解为多个较为简单的子任务,通过增加任务难度帮助智能体逐步适应复杂任务,能够避免过早引入复杂任务所导致的训练不稳定,从而加速学习过程并提高其稳定性。

本文参考课程学习增加任务难度的特性,考虑不同任务之间的复杂度差异,将任务划分为两个阶段:第 1 阶段,训练无人机保持稳定高度并完成基础的跟踪任务,即无人车始终在视野范围内;第 2 阶段,基于第一阶段的成果,在考虑位置、速度、姿态等多种因素的基础上,进行更高层次的飞行控制任务,从而提升系统的整体性能。若无人机在第 1 阶段中连续完成 20 个回合的稳定任务,即连续 20 回合 timesteps>1024,则进入第 2 阶段。

1.5 计算资源损耗优化情况

计算资源消耗情况主要体现在算法的时间和空间复杂度,时间复杂度由前向传播、反向传播、更新步骤和经验回放池大小所决定,空间复杂度则由神经网络参数、经验回放池和梯度存储决定。由于加入事件触发机制和课程学习只影响更新步骤和训练步数,在此仅分析时间复杂度。

在 DDPG 算法中,假设神经网络的层数为 L,第 i 层的神经元数为 M_i,则单次前向传播的时间复杂度为 $O(C_f) = O(\sum_{i=1}^{L-1} M_i \cdot M_{i+1})$;反向传播涉及计算梯度,复杂度为 $O(C_b) = O(\sum_{i=1}^{L-1} M_i \cdot M_{i+1})$;更新网络参数的时间复杂度与总权重数量 W_{params} 成正比,更新复杂度为 $O(W_{params})$;每个训练步骤从经验回放池中采样数据,假设回放池 R 的大小为 $\mid R \mid$,采样批量大小为 $N \ll \mid R \mid$,则采样复杂度为 $O(N)$。总体时间复杂度取决于训练总步数 T,为 $O(T \cdot N \cdot (C_f + C_b) + T \cdot W_{params})$。

设状态变化触发的概率为 $P_{trigger}$(状态满足触发条件的比例),则仅状态触发后进行完整更新。假设任务划分为 K 个阶段,第 i 阶段的训练步数为 T_i,则总训练步数为 $T = \sum_{i=1}^{K} T_i$,各阶段训练步数随难度提升而提升,合理划分阶段能够降低总训练步数。综上,结合事件触发机制和课程学习的总体时间复杂度可以用式(7)表示。

$$O(T \cdot P_{trigger} \cdot N \cdot (C_f + C_b) + T \cdot P_{trigger} \cdot W_{params}) + O(T \cdot (1 - P_{trigger})) \tag{7}$$

其中，第一项是更新动作时的复杂度，考虑了前向传播以及反向传播的计算；第二项是状态未触发时的复杂度，仅需直接复用动作，计算量可近似忽略。相较于原始 DDPG 算法，触发频率越低（即 $P_{trigger} \to 0$），任务阶段划分越合理，计算资源损耗越小。

2　基于 ETC-DDPG 的无人机跟踪算法设计

2.1　状态空间和动作空间

由于本文是基于 YOLO 进行无人机跟踪，主要信息来源于深度相机传达的二维视觉图像以及无人机自身的 IMU 传感器，所以将观测空间设置为 6 个维度，如式(8)所示。

$$\boldsymbol{S} = (d_x, d_y, v_x, v_y, a_x, a_y) \tag{8}$$

其中，d_x 和 d_y 分别为无人机与目标在 x 轴和 y 轴方向上的距离，这提供了目标相对于无人机的水平和垂直位置信息；v_x 和 v_y 分别为无人机在 x 轴和 y 轴方向上的速度；a_x 和 a_y 分别为无人机在 x 轴和 y 轴方向上的加速度。为了消除量纲影响，以上观测值都归一化到[−1，1]区间。

在无人机控制系统中，姿态变化会直接影响无人机的推进方向、加速度和空气的相互作用，横滚角、俯仰角和偏航角通过控制无人机的各个方向上的受力改变其速度。由于本文的环境是三维空间，所以将动作空间设置为三维的连续空间，如式(9)所示。

$$\boldsymbol{A} = (\varphi, \theta, \phi) \tag{9}$$

其中，$\varphi \in [-10, 10]$ 为偏航角，$\theta \in [-3, 3]$ 为俯仰角，$\phi \in [-3, 3]$ 为横滚角(角度采用弧度制表示)。

2.2　奖励函数

本文任务是让无人机追踪无人车，使无人车保持在视野中心的同时保持自身飞行稳定性，避免出现危险行为。设置距离误差 $e_{distance}$、加速度误差 $e_{acceleration}$、速度误差 $e_{velocity}$ 和动作误差 e_{action} 如式(10)所示。

$$\begin{cases} e_{distance} = \dfrac{|d_x|}{\max d_x} + \dfrac{|d_y|}{\max d_y} \\ e_{acceleration} = \left(\dfrac{|a_x|}{\max a_x}\right)^2 + \left(\dfrac{|a_y|}{\max a_y}\right)^2 \\ e_{velocity} = \mathrm{clip}\left(\dfrac{\max|v_x|}{\max v_x}, 0, 1\right)^2 + \min\left(\left(\dfrac{\min|v_y|}{\max v_y}\right)^2, 1\right) \\ e_{action} = \dfrac{|\varphi|}{\max\varphi} + \dfrac{|\theta|}{\max\theta} + \dfrac{|\phi|}{\max\phi} \end{cases} \tag{10}$$

文献[14]中的奖励函数设计方式是将所有奖励设置为负奖励，无人机与目标的距离越小、飞行越平滑稳定惩罚越小，如式(11)所示。

$$R_{total} = -w_1 \cdot e_{distance} - w_2 \cdot e_{acceleration} - w_3 \cdot e_{velocity} - w_4 \cdot e_{action} \tag{11}$$

其中，$w_i \geqslant 0, i = 1, 2, 3, 4$。负奖励(惩罚)的优势在于为智能体提供直接反馈，帮助其避免不良行为。然而，过度惩罚可能导致智能体过于谨慎，偏向于避免惩罚的策略，而非最优策略，从而降低学习效率并影响最终性能。

为了克服这些弊端，本文对奖励函数进行改进，增加了正奖励机制。该机制不仅惩罚不良行为，也奖励良好行为，从而激励智能体探索更多可能的行动方案，并引导其学习更优策略，奖励函数设计如下：

$$R_{total} = r_{distance} + r_{acceleration} + r_{velocity} + r_{action} + r_{stable\_flight} + r_{done} \tag{12}$$

$r_{distance}$ 是距离奖励，当距离误差小于 0.1 时额外给予奖励，当距离误差大于 0.9 时则额外施加边缘惩罚。

$$r_{distance} = \begin{cases} -\omega_{distance} \cdot e_{distance} + 50, & e_{distance} < 0.1 \\ -\omega_{distance} \cdot e_{distance} - 200, & e_{distance} > 0.9 \\ -\omega_{distance} \cdot e_{distance}, & \text{其他} \end{cases} \tag{13}$$

$r_{acceleration}$ 和 $r_{velocity}$ 是加速度和速度奖励，引导无人机放慢飞行速度。

$$r_{acceleration} = -\omega_{acceleration} \cdot e_{acceleration} \tag{14}$$

$$r_{velocity} = -\omega_{velocity} \cdot e_{velocity} \tag{15}$$

r_{action} 和 r_{stable_flight} 是动作奖励，惩罚大的滚转和俯仰角度，鼓励无人机调整飞行姿态，防止危险飞行导致坠机。

$$r_{action} = -\omega_{action} \cdot e_{action} \tag{16}$$

$$r_{stable\_flight} = 20, \quad e_{distance} < 0.1 \text{ 或 } e_{action} < 0.05 \tag{17}$$

其中，$\omega \geqslant 0$ 为奖励系数。

r_{done} 是视野奖励，只有当无人车在视觉画面中出现的时间超过 1 024 个时间步时给予视觉奖励。

$$r_{done} = 500, \quad timesteps > 1\,024 \tag{18}$$

2.3　算法整体框架

基于事件触发机制及课程强化学习的无人机视觉跟踪，根据无人机的状态空间和动作空间特点，本文设计了 ETC-DDPG 算法，以有效地提高跟踪精度和计算效率。算法结构如图 2 所示，算法实现过程如算法 1 所示。

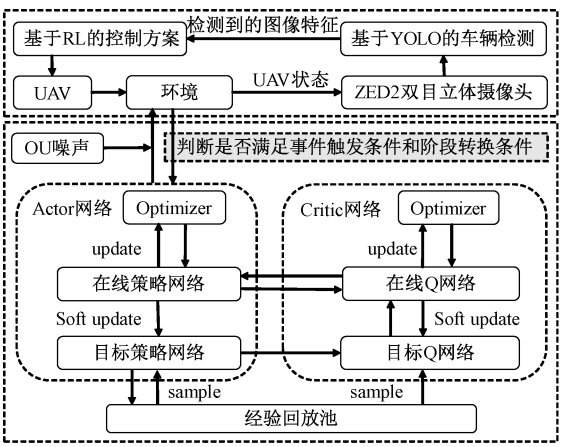


图 2　ETC-DDPG 算法结构图

Fig. 2　Structure diagram of ETC-DDPG algorithm

算法 1:ETC-DDPG 算法

初始化 Critic 网络 $Q(s,a\mid\theta_Q)$ 和 Actor 网络 $\mu(s\mid\theta_\mu)$ 的参数 θ_Q 和 θ_μ

初始化目标网络参数:$\theta'_\mu\leftarrow\theta_\mu$,$\theta'_Q\leftarrow\theta_Q$

初始化经验回放池 R

初始化课程学习阶段 stage=1

for episode $=1,M$ *do*:

随机噪声 $\mathcal{N}_t$ 初始化,获得初始观察状态 s_1

if 满足阶段转换条件

切换课程学习阶段

if 满足事件触发条件 $e(t)>0.1$

执行动作 $a_t=\mu(s_t\mid\theta_\mu)+\mathcal{N}_t$

else 执行动作 $a_t=a_{t-1}$

获取奖励 r_t 和状态 s_{t+1}

将 (s_t,a_t,r_t,s_{t+1}) 存入经验回放池 R

从经验回放池 R 中随机采样批量数据 (s_i,a_i,r_i,s_{i+1})

计算 $y_i=r_i+\gamma Q_{\theta'_Q}(s_{i+1},\mu(s_{i+1}\mid\theta'_\mu))$

更新 critic 参数:最小化损失函数

$$L(\theta_Q)=\frac{1}{N}\sum_i(y_i-Q_{\theta_Q}(s_i,a_i))^2$$

更新 actor 参数:

$$\nabla_{\theta_\mu}J(\theta_\mu)==\frac{1}{N}\sum_{i=1}^{N}\nabla_{\theta_\mu}\mu(s_i\mid\theta_\mu)\nabla_a Q_{\theta_Q}(s_i,a_i)\mid_{a_i=\mu(s_i\mid\theta_\mu)}$$

更新目标网络:$\theta'_\mu\leftarrow\tau\theta_\mu+(1-\tau)\theta'_\mu$,$\theta'_Q\leftarrow\tau\theta_Q+(1-\tau)\theta'_Q$

end for

3 实验及分析

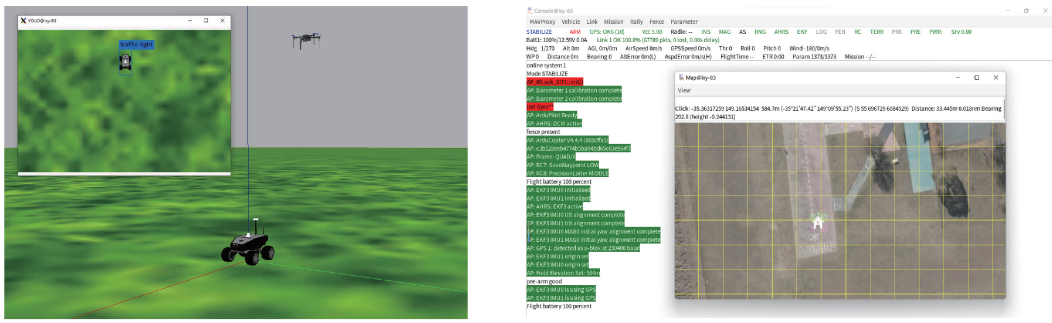
基于文献[14]对无人机跟踪任务的研究,为了提升无人机跟踪性能,本文在改进文献[14]的奖励函数的基础上提出了结合事件触发机制和课程学习的 ETC-DDPG 算法,通过一系列对比和消融实验,验证了 ETC-DDPG 算法的有效性。在鲁棒性测试中,通过调整噪声参数、模型权重和神经元数量,证明该算法良好的鲁棒性和稳定性。

3.1 实验环境

本文基于机器人操作系统(ROS)和 Gazebo 仿真框架构建了仿真环境。该环境部署了 3DR Iris 四旋翼无人机和 Summit XL 无人车,其中,无人机和无人车均配备了传感器套件,包括 GPS、IMU(惯性测量单元)、气压计等,以模拟真实世界中的传感器功能。此外,无人机搭载了 ZED2 下视立体摄像系统,能够提供基于帧的图像数据,可以实时获取图像数据,为仿真环境中的物体识别和追踪任务提供了重要的感知支持。为了实现 ROS 与仿真环境中无人机的通信,本文采用了 MAVLink 协议,该协议支持 ROS 节点与 ArduPilot 飞行控制系统的数据交换,允许在仿真中使用 ArduPilot 的控制算法和传感器数据处理功能。实验环境如图 3 所示,图 3(a)为 SITL 的 Gazebo 环境界面,图 3(b)为 MAVProxy 的地面站画面。实验各项参数如表 1 所示。

3.2 实验结果

在仿真实验中,无人机的初始位置设置为(0, 0, 4),初始航向角为 90°,无人车的位置为(0, 0, 0),每当无人车离



(a) Gazebo仿真环境及YOLO识别画面
(a) Gazebo simulation environment and YOLO recognition screen

(b) MAVProxy地面站画面
(b) MAVProxy ground station screen

图 3 实验环境

Fig. 3 Environment of simulation experiment

开无人机视野,无人机则回到初始位置开始新回合。对于距离误差和偏航角度变化情况,由于目的是使误差收敛至 0,故将误差和角度限制在[-1, 1]之间。在成功率验证方面,考虑到无人机无法完全提前了解目标的真实轨迹以及实际应用场景中目标的大小和所需的操作精度,当距离误差 $e_{\text{distance}}<0.2$ 时即视为成功跟踪,利用成功跟踪的时间步在总时间步中的占比作为成功率。成功率验证实验中设置了静止无人车和随机行驶无人车(行驶方向、速度随机)作为无人机的跟踪对象。此外,本文还设置算法决策时间和轨迹抖动性作为跟踪效率的指标,其中,算法决策时间是指从目标检测到控制命令完成的时间间隔,能够反映策略响应速度,并且在有限硬件性能下还能够衡量不同算法的计算资源需求情况;轨迹抖动性指加速度随时间的变化率,衡量运动轨迹的突然变化,值越小表示运动越平滑。算法训练稳定性则通过训练收敛率和鲁棒性测试验证。算法训练收敛结果如图 4 所示,算法各项指标如表 2 所示。

表 1　参数设置

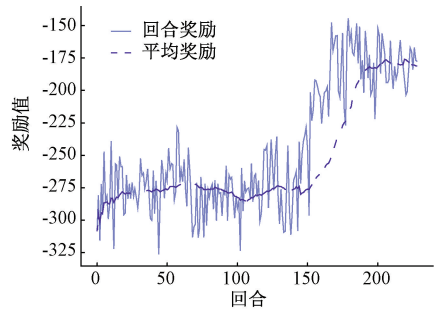
Table 1　Parameter setting

参数名称	参数值
折扣因子 γ	0.99
价值网络学习率 α_Q	0.001
策略网络学习率 α_μ	0.000 1
目标网络参数更新速度 τ	0.001
Replay Buffer	100 000
Mini-batch Size	64
OU 噪声标准差 σ_1	0.1
OU 噪声的衰减速率 σ_2	0.15
每回合最大时间步	1 024

分析图 4，图 4(a)中 ETC-DDPG 算法在 40 000 时间步(170 回合)开始收敛，随着算法训练，图 4(b)显示距离误差逐渐收敛于 0。表 2 所示，ETC-DDPG 算法在静态小车跟踪任务中实现了 99.298%的成功率，在动态小车跟踪任务中仅表现出 5.941%的性能下降，保持了 93.357%的高成功率。这一结果表明 ETC-DDPG 算法在处理动态目标时展现出显著优越的跟踪性能。

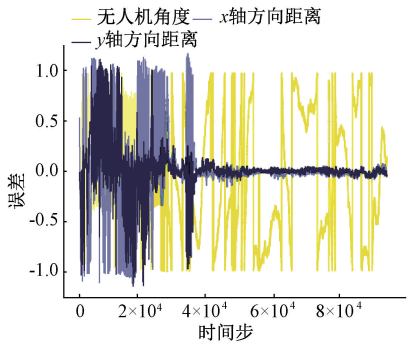
3.3　对比实验

为了对比 DDPG 与 ETC-DDPG 的效果，本文复现文献[14]的 DDPG 算法，使用该文献的奖励函数，并以此作为对比基准进行性能评估。为避免算法名称混乱，本文将其标注为原始 DDPG 算法，算法训练收敛结果如图 5 所示。



(a) ETC-DDPG算法收敛情况

(a) Convergence of ETC-DDPG algorithm



(b) 训练时误差情况

(b) Error during training

图 4　ETC-DDPG 算法训练结果

Fig. 4　Training results of ETC-DDPG algorithm

表 2　算法稳定性、跟踪效率和计算资源需求情况

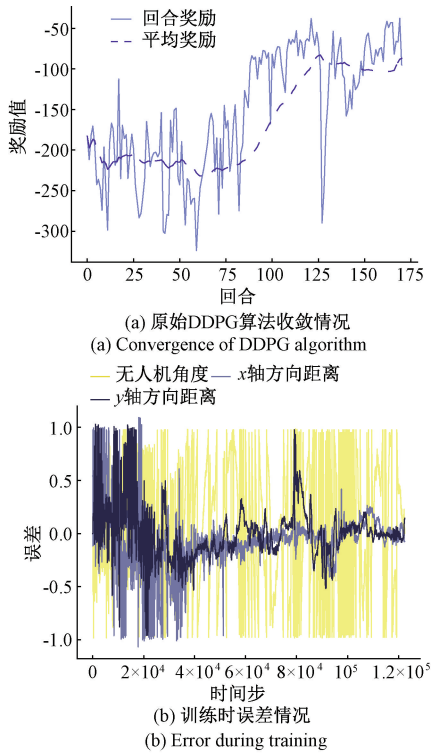
Table 2　Comparison of algorithm stability, tracking efficiency and computing resource consumption

算法	算法奖励函数使用情况	训练收敛率/%	静态小车成功率/%	动态小车			
				成功率/%	平均跟踪误差	算法决策时间/s	轨迹抖动性
original DDPG	文献[14]	75	98.554	37.182	0.309 883 53	0.165	0.016 412 78
DDPG	本文	5	99.410	63.335	0.187 829 81	0.158	0.006 238 33
ET-DDPG	本文	10	99.375	77.742	0.135 017 51	0.031	0.005 416 71
C-DDPG	本文	98	99.565	81.364	0.123 047 78	0.153	0.007 608 95
ETC-TD3	文献[14]	25	98.400	56.257	0.165 246 49	0.155	**0.004 245 52**
ETC-DDPG	本文	**100**	**99.298**	**93.357**	**0.078 959 28**	**0.023**	0.004 431 43

分析图 5(a)，算法在 50 000 时间步(90 回合)时能够收敛，但飞行不稳定，会出现奖励大幅度下跌的情况。相比图 4(b)，图 5(b)的偏航角度变化更杂乱，距离误差明显增大并且中途有大幅波动，证明原始 DDPG 算法收敛速度更慢、飞行效果更不稳定。如表 2 所示，在静态小车跟踪任务中，原始 DDPG 算法展现了高达 98.554%的跟踪成功率。然而，当目标转变为动态小车时，其成功率显著下降至 37.182%，相比 ETC-DDPG 算法平均跟踪误差率上升 292.460%，决策时间增加 0.142 s。

除了 DDPG 算法，在动态目标视觉跟踪领域中最新研究的算法还有柔性动作-评价(soft actor-critic，SAC)、TD3 和近端策略优化(proximal policy optimization，PPO)。由于 SAC 引入了熵最大化，自适应探索能力强，在具有动态障碍物等干扰的复杂环境表现更佳，而 PPO 计算开销大，更适合资源充足的硬件。TD3 在继承 DDPG 硬件友好性的同时，通过双重 Critic 网络和延迟策略更新减少 Q 值过估计问题，因此本文选择 TD3 算法和引入事件触发机制和课程学习的 ETC-TD3 算法进行对比。由于 ETC-DDPG 中的改进奖励函数复杂度上升，算法收敛困难，故本文采用文献[14]中的奖励函数进行算法训练，训练结果如图 6 所示。

分析图 6(a)，TD3 算法在训练过程中陷入了局部最优。图 6(c)显示算法在 70 000 时间步(180 回合)呈现收敛趋势，但收敛过程出现大幅误差的情况。图 6(b)和(d)显示无人机的偏航角度变化剧烈问题。表 2 所示，ETC-



(a) 原始DDPG算法收敛情况
(a) Convergence of DDPG algorithm

(b) 训练时误差情况
(b) Error during training

图 5　原始 DDPG 算法训练结果

Fig. 5　Training results of original DDPG algorithm

TD3 算法跟踪动态目标的成功率只有 56.257%。

以上实验结果表明,相比原始 DDPG 算法和 ETC-TD3 算法,ETC-DDPG 算法显著提升了无人机动态目标跟踪任务的性能,使其在保持快速收敛的同时,实现更稳定的飞行控制和更好的跟踪动态目标性能。

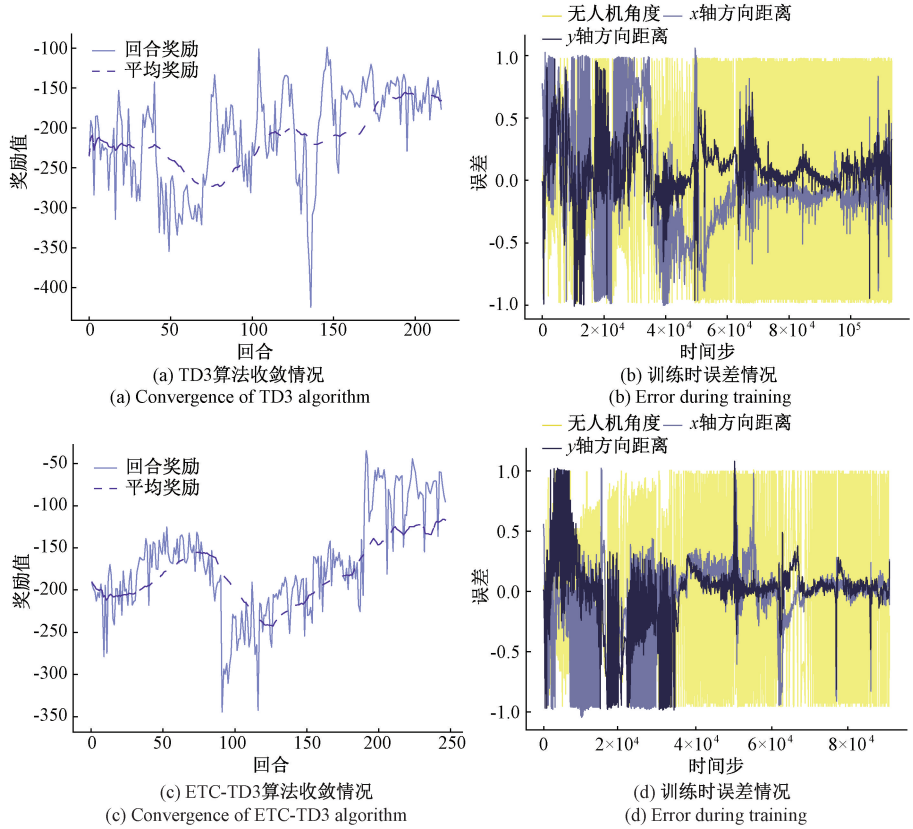
3.4　消融实验

为了系统地评估 ETC-DDPG 算法中各个组成部分对训练性能的贡献和影响,本文对比了在 ETC-DDPG 算法基础上去除事件触发机制和课程学习的 DDPG 算法、只去除课程学习的 ET-DDPG 算法、只去除事件触发机制的 C-DDPG 算法,各算法的训练结果如图 7~9 所示。

分析图 7(a),在去除事件触发机制和课程学习后,相比于 ETC-DDPG 算法,DDPG 算法收敛所需时间变长,需要 70 000 时间步(180 回合)才能收敛,并且图 7(b)显示收敛后无人机跟踪精度较差,误差波动较大。

分析图 8(a),与 ETC-DDPG 算法相比,ET-DDPG 算法收敛速度同样变慢,需要 70 000 时间步(125 回合)收敛。但与 DDPG 算法相比,图 8(b)中 ET-DDPG 算法跟踪误差显著降低,飞行更稳定,并且收敛所需回合数减少,这表明事件触发机制使无人机的探索效率得到提高。

分析图 9(a),C-DDPG 算法在 50 000 时间步(90 回合)收敛,与 ETC-DDPG 算法相比收敛速度相同,但图 9(b)显示偏航角度变化剧烈,说明无人机飞行稳定性较



(a) TD3算法收敛情况
(a) Convergence of TD3 algorithm

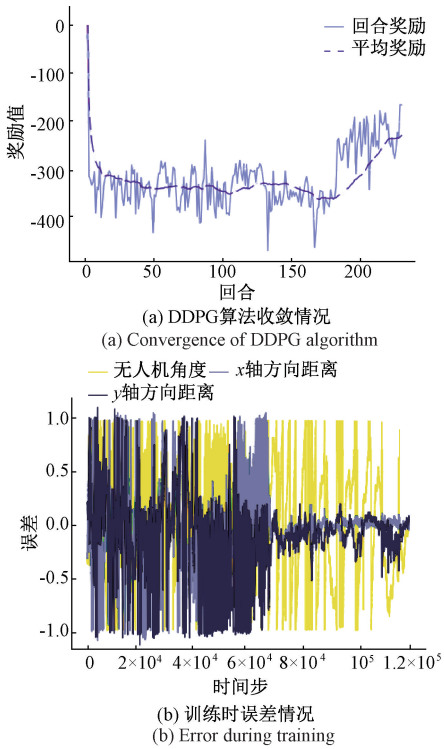
(b) 训练时误差情况
(b) Error during training

(c) ETC-TD3算法收敛情况
(c) Convergence of ETC-TD3 algorithm

(d) 训练时误差情况
(d) Error during training

图 6　TD3 和 ETC-TD3 算法训练结果

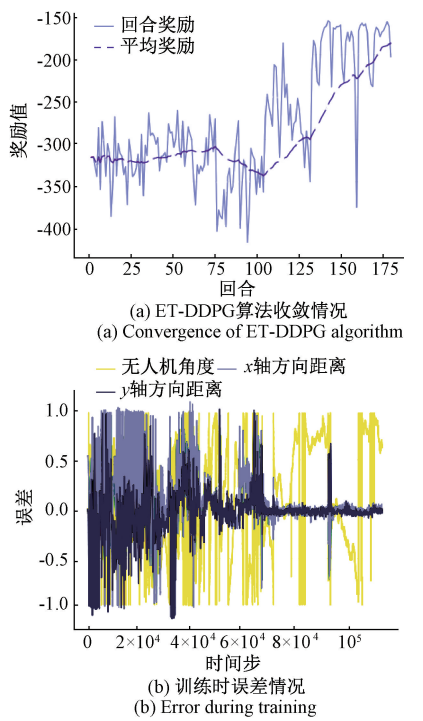
Fig. 6　Training results of TD3 and ETC-TD3 algorithm



(a) DDPG算法收敛情况
(a) Convergence of DDPG algorithm
(b) 训练时误差情况
(b) Error during training

图 7　DDPG 算法训练结果

Fig. 7　Training results of DDPG algorithm

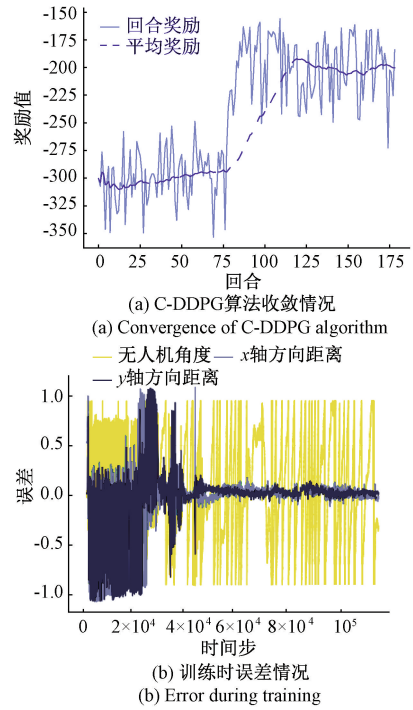


(a) ET-DDPG算法收敛情况
(a) Convergence of ET-DDPG algorithm
(b) 训练时误差情况
(b) Error during training

图 8　ET-DDPG 算法训练结果

Fig. 8　Training results of ET-DDPG algorithm

差。对比 DDPG 和 ET-DDPG，课程强化学习使训练速度得到提升，提高了无人机对复杂任务的适应能力和学习效率。



(a) C-DDPG算法收敛情况
(a) Convergence of C-DDPG algorithm
(b) 训练时误差情况
(b) Error during training

图 9　C-DDPG 算法训练结果

Fig. 9　Training results of C-DDPG algorithm

验证上述算法稳定性、跟踪效率和计算资源需求情况，得到算法各项指标汇总如表 2 所示。

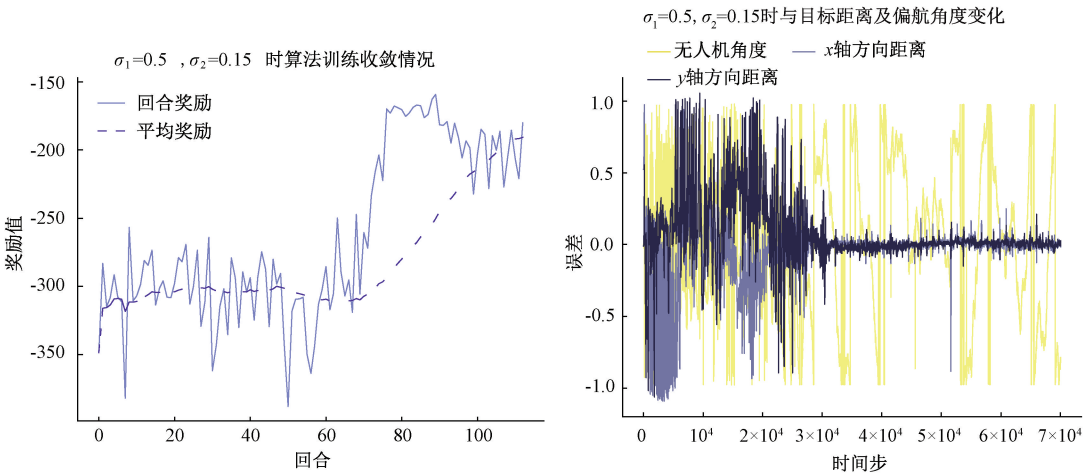
综合分析显示，各算法训练所得模型在静态目标跟踪任务中均实现了高成功率，引入事件触发机制和课程学习策略则显著提升了无人机对动态小车的跟踪性能，事件触发机制能够降低策略响应时间、减少计算资源需求，课程学习能够加快训练速度、提升训练收敛率，而两者的结合则能够在此基础上进一步优化跟踪效果，减小跟踪误差和轨迹抖动性。

3.5　鲁棒性实验

为了验证算法鲁棒性，本文调整了算法超参数。噪声方差 σ_1 和噪声衰减速率 σ_2 会影响智能体的探索，增大方差和减小噪声衰减速率可能导致智能体过度探索，难以稳定优化策略，因此本文将噪声方差增大为 0.5，衰减速率减小为 0.1。改变噪声参数后的算法收敛情况如图 10 所示。

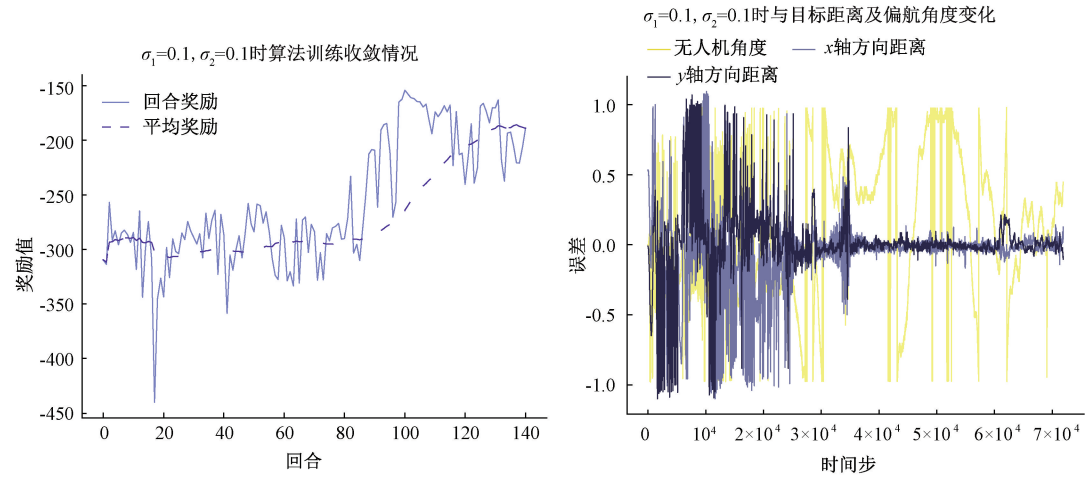
不同的权重初始化方法会影响模型的收敛速度和稳定性。在 ETC-DDPG 算法中，策略网络使用从 −0.003～0.003 的均匀分布初始化权重，价值网络使用 Xavier 初始化。将策略网络的权重初始化方式更改为 He 初始化，同时，策略网络和价值网络的隐藏层神经元数量均增加至原来的两倍。改变模型权重和神经元数量的收敛情况如图 11 所示。

改变噪声参数、权重初始化方式和神经元数量的情况下，ETC-DDPG 算法依旧能够实现稳定收敛，体现了算法



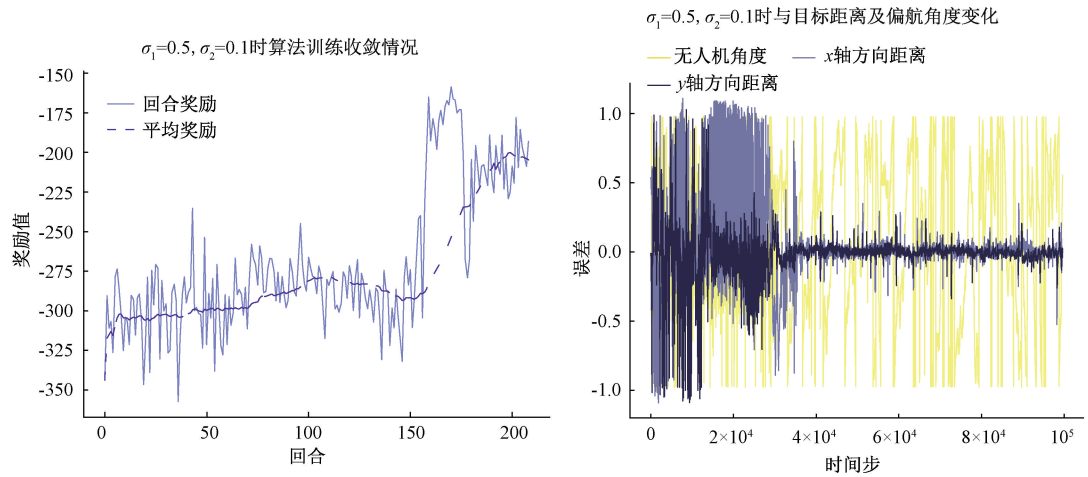
(a) σ_1=0.5, σ_2=0.15时算法收敛情况和训练误差情况

(a) Algorithm convergence and training error of σ_1=0.5, σ_2=0.15



(b) σ_1=0.1, σ_2=0.1时算法收敛情况和训练误差情况

(b) Algorithm convergence and training error of σ_1=0.1, σ_2=0.1

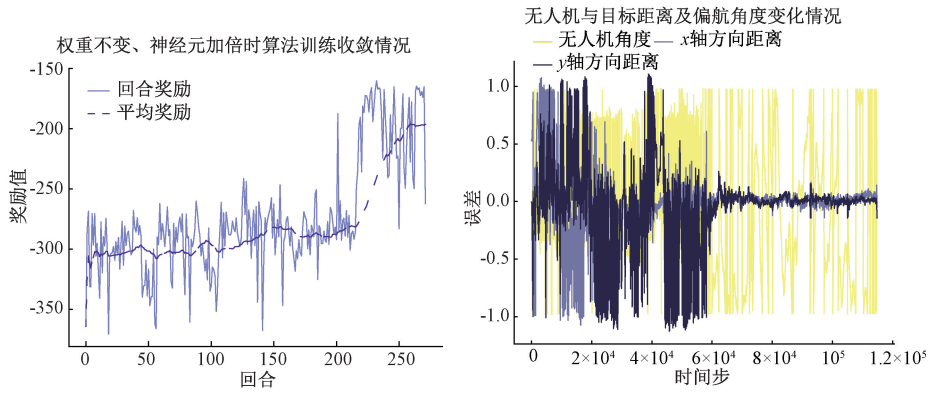


(c) σ_1=0.5, σ_2=0.1时算法收敛情况和训练误差情况

(c) Algorithm convergence and training error of σ_1=0.5, σ_2=0.1

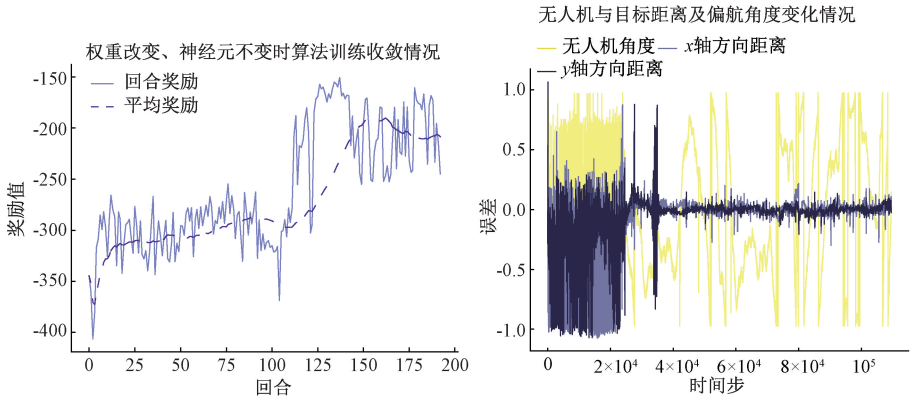
图 10 不同噪声参数收敛情况

Fig. 10 Convergence of different noise parameters



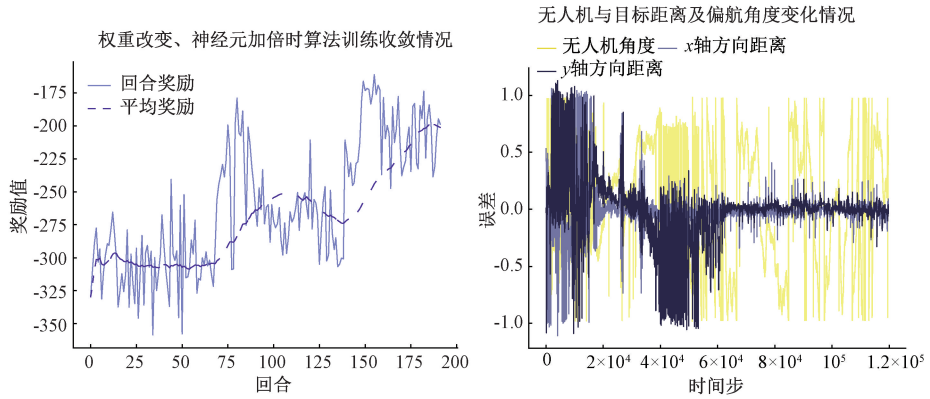
(a) 权重不变、神经元加倍算法收敛情况和训练误差情况

(a) Algorithm convergence and training error with unchanged weights and doubled neurons



(b) 权重改变、神经元不变算法收敛情况和训练误差情况

(b) Algorithm convergence and training error with changed weights and unchanged neurons



(c) 权重改变、神经元加倍算法收敛情况和训练误差情况

(c) Algorithm convergence and training error with changed weights and doubled neurons

图 11　不同权重初始化方式和神经元数量收敛情况

Fig. 11　Convergence of different weight initialization methods and neuron counts

具有鲁棒性，证明了算法的良好训练平稳性。

4　结　　论

本文针对 DDPG 算法在无人机跟踪任务中的计算资源损耗大、收敛速度慢等问题，提出结合事件触发机制和课程学习的 ETC-DDPG 算法。实验结果证明事件触发机制有效提高飞行稳定性和探索效率、减少计算资源需求、降低计算资源损耗，课程学习则提高了算法的收敛速度，二者优势互补，其协同效应进一步增强了算法的鲁棒性和跟踪效率。但本研究主要针对单一目标跟踪任务，未充分考虑多目标、多障碍物等复杂场景。在实际应用中，无人机可能需要同时处理多个目标和复杂环境干扰。后续工作将研究无人机与地面车辆在多障碍物环境中的空地协同，进一步提升系统的复杂环境适应能力。

参考文献

[1] YANG D CH, WANG J, WU F H, et al. Energy efficient transmission strategy for mobile edge computing network in UAV-based patrol inspection system[J]. IEEE Transactions on Mobile Computing, 2024, 23(5): 5984-5998.

[2] 郭嘉琪, 景超, 李雪薇, 等. 融合单目深度和 RTK 定位的电力线弧垂测量方法[J]. 电子测量技术, 2024, 47(2): 89-97.
GUO J Q, JING CH, LI X W, et al. Integration of monocular depth and RTK localization for electric power line sag measurement method[J]. Electronic Measurement Technology, 2024, 47(2): 89-97.

[3] 黄郑, 谢彧颖, 张欣, 等. 基于运动预测与改进 APF 的无人机路径规划方法[J]. 电子测量技术, 2023, 46(24): 103-111.
HUANG ZH, XIE Y Y, ZHANG X, et al. Unmanned aerial vehicle path planning method based on motion prediction and enhanced APF[J]. Electronic Measurement Technology, 2023, 46(24): 103-111.

[4] 何飞麒. 四旋翼无人机俯拍视角下的行人检测与轨迹追踪[J]. 电子测量技术, 2022, 45(10): 50-56.
HE F Q. Pedestrian detection and route tracking from aerial view of quad-rotor UAVs [J]. Electronic Measurement Technology, 2022, 45(10): 50-56.

[5] ALHAFNAWI M, BANY SALAMEH H A, MASADEH A, et al. A survey of indoor and outdoor UAV-based target tracking systems: Current status, challenges, technologies, and future directions[J]. IEEE Access, 2023, 11: 68324-68339.

[6] WILSON A N, KUMAR A, JHA A, et al. Embedded sensors, communication technologies, computing platforms and machine learning for UAVs: A review[J]. IEEE Sensors Journal, 2022, 22(3): 1807-1826.

[7] LI M X, ZHAI D H, YANG D, et al. BVTracker: Multivehicle tracking based on behavioral-visual features[J]. IEEE Sensors Journal, 2023, 23(11): 11815-11824.

[8] QIAN Y ZH, WANG Z J, GAO Y W, et al. A multi-object tracking method in moving UAV based on IoU matching[J]. IEEE Access, 2024, 12: 139076-139085.

[9] SUN N Y, ZHAO J, SHI Q, et al. Moving target tracking by unmanned aerial vehicle: A survey and taxonomy [J]. IEEE Transactions on Industrial Informatics, 2024, 20(5): 7056-7068.

[10] ZHU P F, WEN L Y, DU D W, et al. Detection and tracking meet drones challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7380-7399.

[11] JAGATHEESAPERUMAL S K, HASSAN M M, HASSAN Md R, et al. The duo of visual servoing and deep learning-based methods for situation-aware disaster management: A comprehensive review[J]. Cognitive Computation, 2024, 16(5): 2756-2778.

[12] 华夏, 王新晴, 芮挺, 等. 视觉感知的无人机端到端目标跟踪控制技术[J]. 浙江大学学报(工学版), 2022, 56(7): 1464-1472.
HUA X, WANG X Q, RUI T, et al. Vision-driven end-to-end maneuvering object tracking of UAV[J]. Journal of Zhejiang University(Engineering Science), 2022, 56(7):1464-1472.

[13] WANG CH, WANG J Q, MA ZH W, et al. Integrated learning-based framework for autonomous quadrotor UAV landing on a collaborative Moving UGV[J]. IEEE Transactions on Vehicular Technology, 2024, 73(11): 16092-16107.

[14] MITAKIDIS A, ASPRAGKATHOS S N, PANETSOS F, et al. A deep reinforcement learning visual servoing control strategy for target tracking using a multirotor UAV[C]. 2023 9th International Conference on Automation, Robotics and Applications (ICARA), 2023: 219-224.

[15] WANG X, CHEN Y D, ZHU W W. A survey on curriculum learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(9): 4555-4576.

[16] 马澳华, 邢关生. 基于 GRU-A3C 的四旋翼无人机视觉避障系统[J]. 电子测量技术, 2024, 47(21): 46-52.
MA AO H, XING G SH. Visual obstacle avoidance system for quadrotor UAV based on GRU-A3C[J]. Electronic Measurement Technology, 2024, 47(21): 46-52.

[17] HU Z J, GAO X G, WAN K F, et al. Asynchronous curriculum experience replay: A deep reinforcement learning approach for UAV autonomous motion control in unknown dynamic environments[J]. IEEE Transactions on Vehicular Technology, 2023, 72(11): 13985-14001.

[18] SEDGHI L, IJAZ Z, NOOR-A-RAHIM M, et al. Machine learning in event-triggered control: Recent advances and open issues[J]. IEEE Access, 2022, 10: 74671-74690.

[19] 徐鹏, 谢广明, 文家燕, 等. 事件驱动的强化学习多智能体编队控制[J]. 智能系统学报, 2019, 14(1): 93-98.

XU P, XIE G M, WEN J Y, et al. Event-triggered reinforcement learning formation control for multi-agent[J]. CAAI Transactions on Intelligent Systems, 2019, 14(1): 93-98.

[20] TERVEN J, CORDOVA-ESPARZA D. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.

[21] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. ArXiv preprint arXiv:1509.02971, 2019.

作者简介

庄晶颖，硕士研究生，主要研究方向为强化学习。

E-mail:zjingying0905@163.com

刘磊（通信作者），博士，教授，博士生导师，主要研究方向为无人系统和强化学习。

E-mail:liulei_hust@163.com

闫冬梅，博士，讲师，主要研究方向为强化学习、交通网络均衡分配和控制。

E-mail:ydm@njupt.edu.cn

梁成庆，博士研究生，主要研究方向为多智能体强化学习、无人机集群编队。

E-mail:lchengq_hhu@163.com