https://doi.org/10.1051/jnwpu/20264410001

基于深度强化学习的中小型无人机高原峡谷抗风飞行控制

朱越, 王睿, 周洲

(西北工业大学 航空学院, 陕西 西安 710072)

摘 要:固定翼无人机具有续航时间长、距离远等特点,在高原峡谷等复杂地形区域的大范围巡察任务中更具优势。然而该类地形常伴随强烈多变的风场,严重威胁无人机的飞行安全与轨迹稳定性。针对以上问题,开展了面向典型高原峡谷风场环境的深度强化学习(DRL)横航向轨迹控制研究,提出一种以 L1 制导律为基准的补偿式 DRL 控制策略。以控制侧向轨迹为目标设计评价函数,基于简化动力学模型与保留峡谷风场特征的简化风场模型开展策略训练,在保证建模精度的同时实现快速训练,所训练策略成功迁移至六自由度高保真模型及半物理仿真试验平台进行验证。试验结果表明,所提轨迹控制策略对高原峡谷风场扰动具有良好的抑制效果:在最大侧向风速达 16 m/s 的扰动环境中,其轨迹偏差仅为传统 L1 方法的 28.6%,同时展现出良好的迁移特性、鲁棒性与工程可实现性。

关 键 词:深度强化学习;TD3 算法;轨迹控制;极端风场;固定翼无人机
中图分类号:V212.1 **文献标志码**:A **文章编号**:1000-2758(2026)01-0001-11

我国高原地区面积广阔,约占国土总面积的三分之一,该区域具有重要的生态屏障功能与战略价值,但由于高海拔缺氧、地形复杂等恶劣环境特征,传统人工巡查模式面临作业效率低、人员安全风险高、经济成本攀升等多重困境。中小型固定翼无人机具有成本低、机动性强等特点,与旋翼无人机相比还具备显著的航程航时优势,为构建高原广域巡察体系提供了新的解决方案。然而,高原环境地形复杂,风场多变,气流扰动速度可达到无人机巡航速度的 20%~60%,甚至更高。中小型无人机体积小、质量轻、速度低,更易受到大气环境的影响,进而引发姿态不稳定、轨迹偏离等现象,严重时甚至导致无人机失控坠毁。传统控制方法在高动态、非线性和不确定环境下存在一定局限,对精确模型高度依赖;在复杂风场条件下,即使高保真模型亦难以避免建模误差,因此实际控制器需具备基于感知信息的自适应调节能力。目前许多非线性控制方法已应用于风场下的飞行控制设计,包括 H_∞ 控制、非线性动态逆、基于风估计的主动抗风控制、神经元飞行方法等[1-4]。上述研究使无人机闭环系统抗风性能得到显著改善,但通常依赖特定系统假设,且适用范围有限,尚难以有效应对极端环境下的控制需求。

DRL 结合了深度学习(deep learning, DL)和强化学习(reinforcement learning, RL)的优势,通过与环境的持续交互自主学习最优策略,能够有效处理高维状态和动作空间,实现端到端的控制而无需依赖被控对象的准确建模,在无人机抗风控制中展现出良好应用前景。近年来,多项研究将深度强化学习引入旋翼无人机控制与抗风问题,在极端初始条件和复杂风场扰动下均展现出优异性能,并得到了仿真与飞行试验的验证[5-8]。

上述研究结果表明,基于 DRL 的控制算法能够有效提升旋翼无人机的抗风能力。然而相较于旋翼无人机,固定翼无人机的飞行控制难度更高、更易受到风场扰动。在风场中面临更复杂的非定常空气动力学相互作用,该过程的建模难度也显著增加[9]。在此背景下,DRL 方法为复杂风场下无人机飞行提供了一种更具潜力的控制思路,增强系统对复杂扰

收稿日期:2025-04-28
基金项目:国家自然科学基金(12502386)资助
作者简介:朱越(2002—),硕士研究生
通信作者:王睿(1981—),副教授 e-mail:wangrui@nwpu.edu.cn

动的响应能力,同时也减轻对高精度模型的依赖。

然而,固定翼无人机具有飞行条件上的诸多限制,如最小速度、失速迎角等,一旦策略不合理,将直接导致其失控甚至坠毁。在训练初期,旋翼无人机可采用“停止-思考-行动”的方式进行决策以降低探索风险;而固定翼无人机智能体随机行动会导致若干糟糕行为,对这些行为的学习将使 DRL 策略训练更为困难。同时,这一特点也增加了 DRL 策略在实际固定翼飞行器上部署的风险[10]。上述因素提高了 DRL 方法的应用难度,因而迄今为止,基于 DRL 算法的固定翼无人机抗风控制研究较少。

Rennie[11]将飞行动力学模型与 RL 试验相结合,开发了可提供快速配置的飞行控制环境软件包,并基于此开展了一系列评估智能体性能的仿真试验,证实了该算法的有效性,但其在更复杂环境中表现较差;Kong 等[12]在此基础上加入高斯风速变化模型,探究了姿态塑形奖励在可变风速环境中的控制效果,研究证明使用塑形奖励的无人机具有更优的飞行性能;Bohn 等提出了一种 DRL 控制器来处理非线性姿态控制问题,通过数字仿真[13]与实际飞行试验[14]证明了该方法的有效性与鲁棒性,并探讨了如何采用现有控制方法来加速学习型控制器的学习过程[15];Chowdhury 等[16]最先基于 DRL 固定翼无人机控制策略开展不同天气下的实际飞行试验,结果表明,基于 DRL 的控制器展现出更优的跟踪性能,该团队[17]进一步提出一种可移植 DRL 控制器,它能在完全不同的飞机上展现出良好性能,但在飞行试验中却出现高频振荡现象。

迄今为止,基于 DRL 的固定翼无人机控制研究多停留在理论仿真阶段,实际飞行试验较少。仍有一个核心问题尚未解决,即模拟中学习到的策略如何转移到实际飞行中[18]。当前无人机风场控制研究中,大多为某种简单风场或其与紊流的叠加,这对于复杂风场中的飞行控制研究而言远远不够。此外,目前风场环境下的 DRL 控制研究,大多以提升鲁棒性为出发点,将复杂风场下无人机抗风控制与 DRL 策略相结合的针对性研究尚需进一步探索。

针对小型固定翼无人机在高原峡谷风场中的飞行问题,本文根据高原典型峡谷地貌构建了峡谷风场模型,并基于该风场环境以横航向简化线性动力学模型开展 DRL 轨迹控制策略训练,通过设计评价函数、状态空间等手段在保证精度的前提下提高策略收敛速度;进一步将训练得到的轨迹控制器移植到六自由度(6DoF)高保真模型中,通过数字仿真验证了该 DRL 控制策略可行性;最后基于半物理仿真试验平台验证了提出算法的泛化性与工程可实现性,为小型固定翼无人机在极端风场环境下的飞行控制问题提供一种训练代价较低的可行方案。

1 高原峡谷风场模型

本文基于高原常见的峡谷地貌构建了高原峡谷风场模型,能根据峡谷形状和无穷远处的风速风向,快速生成风场。风从外向内吹入该山谷时,会在山谷中央加速,并在山谷的外侧向两边扩散。

1.1 峡谷山体建模

本文将峡谷地形简化,采用与大多山体都较为接近的余弦形山体,山体断面轮廓线表达式为

$$z = H \cdot \cos^2(\pi r/D) \tag{1}$$

考虑一定山脉长度后建立如图 1 所示高原峡谷山体几何模型,其中, D 为山体底部直径,H 为山体高度。

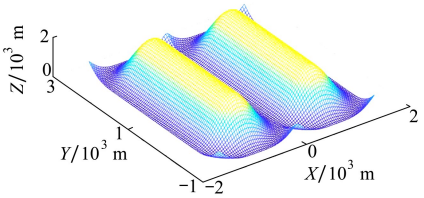


图 1 峡谷地形三维模型

1.2 峡谷风场建模

当气流进入峡谷时,地形对气流压缩作用会产生狭管效应,下文将从风速的垂直分布和水平分布两部分开展讨论。

1.2.1 垂直分布

针对峡谷地形中的山体间距、山体坡度、山体长度等地形因素,峡谷内部平均风速随高度变化趋势可采用文献[19]中的指数函数描述

$$V = V_0 \cdot \mu_W \mu_S \mu_L (Z/Z_0)^{\alpha \cdot \beta_W \beta_S \beta_L} \tag{2}$$

式中:V 为峡谷内部高度为 Z 处的风速;V_0 为参考点风速;Z_0 为参考点高度;α 为地貌粗糙度指数,此处取 0.15;μ_W,μ_S,μ_L,β_W,β_S,β_L 为考虑山体间距、山体坡度、山体长度影响的调整系数。参数的详细取值可见文献[19]。

1.2.2 水平分布

取高度为 Z 处的山体水平截面,如图 2 所示。

其中,实线部分为山体截面,虚线部分为山体底部轮廓。高度 Z 处山体半径为 R,计算公式为

$$R = \arccos\sqrt{Z/H} \cdot D/\pi \tag{3}$$

对于峡谷入口处流场,可以采用叠加均匀流 V_∞ 与峡谷入口两侧对称布置 2 对变强度偶极子的流场来描述。将峡谷入口流场简化为 2 个圆柱绕流问题的叠加,2 对偶极子的强度 K 相同,方向均指向上游,即与均匀来流方向相反,2 对偶极子之间距离为 $2d$。

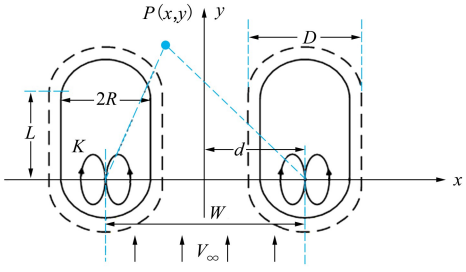


图 2　山体几何模型水平截面

该流场可进一步简化为关于 x-z 和 y-z 平面对称,因而可认为偶极子强度 K 为 Z 的函数 $K(Z)$。流场中任意点 $P(x,y)$ 的诱导速度计算公式为

$$v_x(x,y) = \frac{K}{2\pi}\left(\frac{-2(x+d)y}{((x+d)^2+y^2)^2} + \frac{-2(x-d)y}{((x-d)^2+y^2)^2}\right)$$

$$v_y(x,y) = V_\infty + \frac{K}{2\pi}\left(\frac{(x+d)^2-y^2}{((x+d)^2+y^2)^2} + \frac{(x-d)^2-y^2}{((x-d)^2+y^2)^2}\right) \tag{4}$$

根据圆柱绕流模型公式[20]推导出此时偶极子强度 K 的大小为

$$K = 2\pi V_\infty R^2 \tag{5}$$

结合(3)式即可根据 Z 计算得出 K 的大小。假设远方风速为 14 m/s,山谷底部距离 400 m,则会在图 1 所示高原峡谷中生成如图 3 所示的风场。

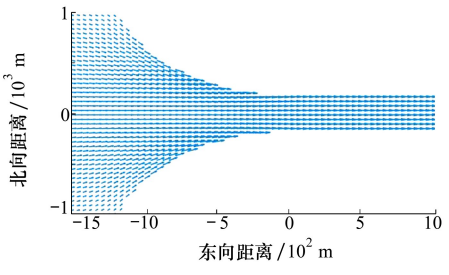


图 3　高原峡谷风场

此外,除了常值风和地形诱导的突风之外,后续仿真中还将加入紊流风以贴合实际风场。

2　风场中的 DRL 横航向轨迹控制

2.1　算例无人机简介

为了便于深入讨论,本文以图 4 所示的一种常见中小型固定翼布局无人机为例。该无人机采用常规式布局、H 形尾翼,由 8 组动力系统驱动。

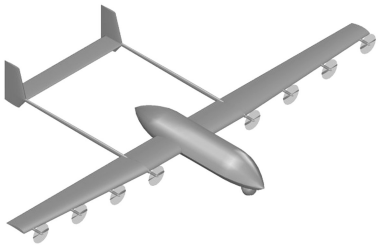


图 4　算例无人机总体布局示意图

2.1.1　无人机的主要参数

算例无人机主要总体布局参数如表 1 所示,其中:m 为无人机质量;S 为无人机参考面积;b 为机翼展长;c 为参考弦长;I_x,I_y,I_z,I_{xz} 为惯量数据。

表 1　算例无人机的总体布局参数

参数	值	参数	值
m/kg	270	I_x/(kg·m²)	90
S/m²	3.6	I_y/(kg·m²)	162
b/m	7.2	I_z/(kg·m²)	235
c/m	0.5	I_{xz}/(kg·m²)	1.8

算例无人机的主要气动数据如表 2 所示。

表 2　算例无人机的主要气动数据

参数	值	参数	值
C_{L0}	0.603	$C_{Y\beta}$	−0.405
$C_{L\alpha}$	6.011	$C_{l\beta}$	−0.079
C_{D0}	0.051	$C_{n\beta}$	0.067
k	0.027	C_{Yp}	−0.093
C_{m0}	0.144	C_{lp}	−0.644
$C_{m\alpha}$	−1.222	C_{np}	−0.057
C_{Lq}	5.189	C_{Yr}	0.290
C_{mq}	−18.482	C_{lr}	0.168

续表 2

参数	值	参数	值
$C_{L\alpha}$	0.321	C_{nr}	- 0.109
$C_{m\alpha}$	3.125	$C_{l\delta_a}$	- 0.290
$C_{L\delta_e}$	0.332	$C_{n\delta_a}$	0.001
$C_{m\delta_e}$	- 1.51	$C_{Y\delta_r}$	0.229
$C_{n\delta_r}$	- 0.075	$C_{l\delta_r}$	0.017
C_{Lmax}	3	$V_{stall}/(\mathrm{m}\cdot\mathrm{s}^{-1})$	25

算例无人机的 8 套动力系统完全一致。每个电机的最大功率 6.25 kW,输出效率 0.9。

2.1.2 飞行控制系统架构

该无人机主要控制构架如图 5 所示。

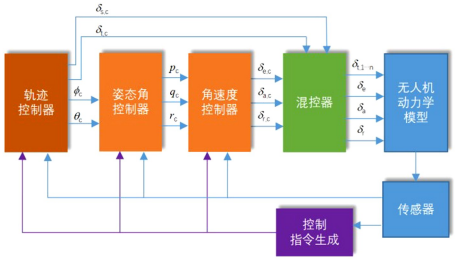


图 5 无人机飞控系统的主要构架

图 5 中,δ_e,δ_a,δ_r 分别为升降舵、副翼、方向舵偏航角;p,q,r 分别为滚转角速度、俯仰角速度、偏航角速度;ϕ_c,θ_c 分别为滚转角和俯仰角的目标值;δ_t 为油门大小;$\delta_{s,c}$ 为侧力控制量。

无人机俯仰、滚转和偏航操纵分别由升降舵、副翼和方向舵执行,角速度控制器和姿态角控制器采用串级 PID 控制方法。轨迹控制器包含纵向和侧向轨迹控制,纵向轨迹控制采用 TECS[21] 方法,侧向轨迹控制包含了 L1 方法[22] 和本文提出的 DRL 方法。本文的 DRL 控制策略以 L1 轨迹控制方法为基准,在原有制导结构前提下,引入 DRL 模块对 L1 制导律进行修正。使用时可根据需要选择 DRL 补偿模块是否发挥作用。

根据 L1 制导律,横向加速度 $a_{s,cmd}$ 为

$$a_{s,cmd} = 2\frac{V^2}{L_1}\sin\eta \tag{6}$$

式中:L_1 为从无人机到目标飞行轨迹上的参考点的线段;η 为从飞行速度到 L_1 线段的夹角。当目标飞行轨迹为沿地轴系 x 轴的直线段时,η 的表达式为

$$\eta = \chi_c - \chi + \sin^{-1}\left(\frac{y_c - y}{L_1}\right) \tag{7}$$

式中,y_c 和 χ_c 为期望轨迹的侧向位移和参考轨迹在地轴系上的方位角;y 和 χ 无人机侧向位移和飞行轨迹在地轴系上的方位角。

本文 DRL 控制策略通过对 L1 制导律输入参数 y_c 和 χ_c 进行补偿抑制风场干扰,常规 L1 控制中,这 2 项由预设的目标轨迹确定。对其进行合理、及时的补偿可有效调节控制律输出,从而提升无人机侧向风场扰动下的响应速度与控制精度。

2.2 基于 DRL 的抗风控制策略

强化学习是一个反复迭代的过程,如图 6 所示,智能体基于感知到的当前状态 s_t,根据策略从动作空间 $\boldsymbol{A}$ 中选择动作 a_t 执行,环境根据 a_t 来反馈相应的奖励 r_t,并依据转移概率 $P(s_{t+1}|s_t,a_t)$ 转移到下一状态 s_{t+1},智能体根据奖励调整自身策略并依据新状态做出决策。在奖励 r 基础上,定义时刻 t 到 T 的累计奖励为回报函数,即

$$R_t = \sum_{i=t}^{T}\gamma^{i-t}r_i \tag{8}$$

式中:γ 为折扣因子,$0 < \gamma \leqslant 1$,表示未来奖励的权重;T 为回合结束时的步长总数。强化学习的目标是使智能体找到最优策略 π^* 使得系统的累计奖励最大化。

以控制量舵偏角作为输出直接开展策略训练时,智能体训练初期的随机探索易出现若干糟糕行为,增加后续策略收敛难度。因此本文提出补偿式 DRL 控制策略,以 L1 控制方法为基准,根据无人机的当前飞行状态信息,输出目标侧偏距 y_c 与目标航向角 χ_c 对原 L1 制导律进行补偿,即本文 DRL 控制策略动作空间为 $\boldsymbol{A} = [y_c, \chi_c]$。

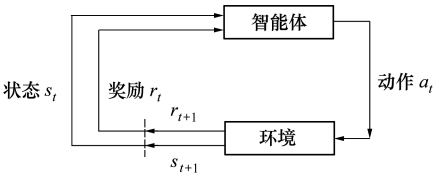


图 6 强化学习基本过程

L1 方法基于飞行器动力学设计,具有理论上的稳定性保证,以此为基准开展算法设计,能缩小策略搜索空间,降低训练初期的不稳定风险。相较于直接输出舵偏角的训练方法,以 y_c 和 χ_c 为输出设计 DRL 策略具有更明确的物理意义与可解释性。

2.2.1　观测量的选择

观测量作为深度神经网络的输入,相当于常规控制中的反馈量,其合理选取是获得良好控制性能的基础。考虑到侧风首先作用于侧滑角 β,进而影响侧向速度 V_y、滚转角速度 p、偏航角速度 r,再进一步影响侧向位移 y、航向角 χ 及相关状态,本文选取 $\beta, V_y, p, r, y, \phi, \chi$ 作为观测量。同时,为减小稳态误差,引入侧向位移偏差积分项 $\int y\mathrm{d}t$。

上述 8 个变量能够较为全面地表征无人机在风场中的运动状态,选择这 8 个变量构成 DRL 控制策略的状态空间,即 $\boldsymbol{S}=\left[\beta, V_y, p, r, y, \phi, \chi, \int y\mathrm{d}t\right]$。

2.2.2　评价函数的设计

评价函数设计是 DRL 算法核心环节,通过将多变量连续系统的状态与动作空间映射为标量奖励信号,主导策略优化方向并决定控制目标与系统性能。

设计评价函数首先需要确定评价变量。本文 DRL 控制策略以侧向轨迹为控制目标,故选择 y 和 V_y 作为奖励量,二者接近于 0 时奖励最大;同时考虑到物理限制,为防止控制量过大、变化过快,引入有关 p、r、目标偏移量导数 $\dot{y}_c$ 和目标航向角导数 $\dot{\chi}_c$ 作为惩罚项。本文设计了如下评价函数:

$$r(x)=r_1(x)+r_2(x)$$

$$r_1(x)=\begin{cases}c_1x^2, & |x|<c_2\\ 2c_1c_2\mathrm{sign}(x)x-c_1c_2^2, & |x|\geqslant c_2\end{cases}$$

$$r_2(x)=c_3\mathrm{e}^{c_4x^2} \tag{9}$$

式中:c_1, c_2, c_3, c_4 分别为决定函数形状的待定系数;x 代表偏差量,0 是其最优值;$r(x)$ 代表对该偏差的评价。可以求得该评价函数的一阶导数为

$$r'(x)=\begin{cases}2c_1x+2c_3c_4x\mathrm{e}^{c_4x^2}, & |x|<c_2\\ 2c_1c_2\mathrm{sign}(x)+2c_3c_4x\mathrm{e}^{c_4x^2}, & |x|\geqslant c_2\end{cases} \tag{10}$$

显然,$\forall x\in\mathbf{R}$,评价函数的一阶导 $r'(x)$ 连续。不失一般性地,假设 $c_1=c_4=-1, c_2=c_3=1$,可以求得 $r(x)$ 的曲线如图 7 所示。

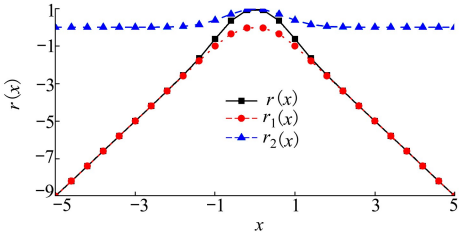


图 7　评价函数曲线图

该评价函数除了一阶导数连续之外,还具有以下优点:①当 x 偏差较大时,函数线性收敛,可以避免多变量系统在大偏差时,由于梯度差距过大不容易收敛到最优解的问题;②当 $x\to0$ 时,梯度 $\to0$,可使 DRL 算法在最优解附近多驻留;③在最优解附近可获得更高奖励,且奖励最大值是可设计的。

本文基于(9)式设计 y 和 V_y 的奖励,并考虑 p, r, $\dot{y}_c$ 和 $\dot{\chi}_c$ 的惩罚,得到第 i 个时间节点 t_i 上的评价函数为

$$r_{t_i}=[w_{t,y}r(y)+w_{t,V_y}r(V_y)-w_{t,p}|p|^2-w_{t,r}|r|^2-w_{t,\dot{y}_c}|\dot{y}_c|-w_{t,\dot{\chi}_c}|\dot{\chi}_c|]T_s/T_f \tag{11}$$

式中:y 为侧向位移偏差;V_y 为侧向速度;$\dot{y}_c$ 为偏移量指令导数;$\dot{\chi}_c$ 为偏航角指令导数;T_s 和 T_f 分别为每个采样步长的时间和每一幕的时间。(11)式中角度单位均为弧度。评价函数系数的取值如表 3 所示。

表 3　评价函数系数的取值

参数	值	参数	值
$c_1\mid r(y)$	−1	$w_{t,y}$	3
$c_2\mid r(y)$	0.5	w_{t,V_y}	0.1
$c_3\mid r(y)$	3	$w_{t,p}$	10
$c_4\mid r(y)$	−3	$w_{t,r}$	10
$c_1\mid r(V_y)$	−1	$w_{t,\dot{y}_c}$	0.01
$c_2\mid r(V_y)$	0.5	$w_{t,\dot{\chi}_c}$	0.4
$c_3\mid r(V_y)$	2	$c_4\mid r(V_y)$	−3

本文以控制侧向轨迹为目标,因而给予与 y 有关的评价值较高的权重。同时,给予与 p, r 有关的惩罚项较大的权重参数,从而迫使智能体选择平滑的姿态调整策略,减少快速或频繁的舵面偏转,使无人机在保证姿态稳定前提下提高轨迹跟踪精度。

2.3　基于 TD3 的轨迹控制器设计

本文采用 TD3 算法进行控制器训练,TD3[23] 是 Twin Delayed Deep Deterministic Policy Gradient 的全称,TD3 算法流程如图 8 所示。

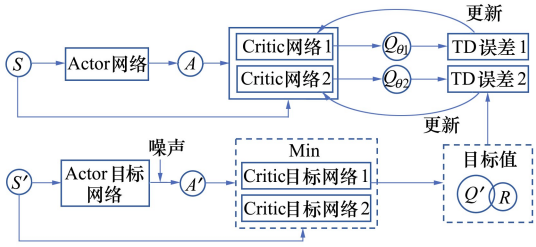


图 8　TD3 算法流程图

首先,使用随机参数 θ_1,θ_2,ϕ 初始化评价网络 $Q_{\theta_1},Q_{\theta_1}$ 和策略网络 π_ϕ,初始化目标网络 $\theta_1'\leftarrow\theta_1,\theta_2'\leftarrow\theta_2,\phi'\leftarrow\phi$ 以及经验缓冲池。对于每一个幕:

1) 初始化 Uhlenbeck-Ornstein(UO) 随机过程,初始化仿真环境并获得无人机的初始状态 s_1;

2) 重复以下过程直至达到最大步长:

(1) 根据策略网络与当前状态 s_t 选择动作 $\pi_\phi(s_t)$,采样 UO 噪声得到 $\grave{o}$,下达动作指令 $a_t=\pi_\phi(s_t)+\grave{o}$ 给无人机模型;

(2) UAV 执行 a_t,返回奖励 r_t 和新状态 s_{t+1};

(3) 储存状态转移信息 (s_t,a_t,r_t,s_{t+1}) 到经验缓冲池,作为训练当前网络的数据集;

(4) 从经验缓冲池抽取 N 个状态转移信息 (s_t,a_t,r_t,s_{t+1}),根据策略网络计算目标动作 $\tilde{a}$,计算 2 个评价网络下的评价值,取较小值作为 y;

(5) 以学习率 α 更新评价网络

$$\theta_i\leftarrow\theta_i-\alpha\cdot\mathrm{argmin}_{\theta_i}N^{-1}\sum(y-Q_{\theta_i}(s,a))^2$$

(6) 延迟更新策略网络,每隔 d 个步长采用确定性策略梯度更新一次策略网络 π_ϕ

$$\nabla_\phi J(\phi)=N^{-1}\sum\nabla_a Q_{\theta_1}(s,a)\Big|_{s=s_t,a=\pi_\phi(s_t)}\nabla_\phi\pi_\phi(s)\Big|_{s=s_t}$$

(7) 更新目标网络

$$\theta_i'\leftarrow\tau\theta_i+(1-\tau)\theta_i'$$
$$\phi'\leftarrow\tau\phi+(1-\tau)\phi'$$

式中,τ 为 polyak 系数,控制每次更新的权重。

3) 根据(8)式计算该幕中所有步长奖励之和。

2.3.1 TD3 算法设计

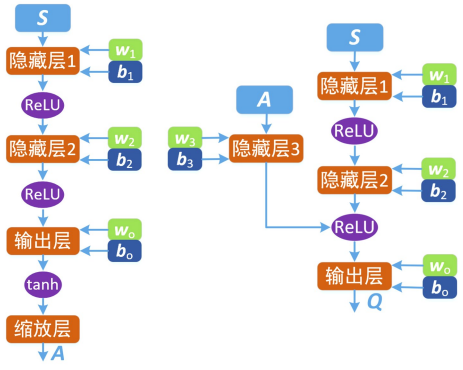
TD3 算法使用了双 Critic 网络及其目标网络,因此加上 Actor 网络及其目标网络,一共使用 6 个深度神经网络。

2.3.2 深度神经网络设计

神经网络的结构参数需在表达能力与计算开销之间权衡,节点数不足将限制对无人机控制-响应的建模能力,而节点数过多则会增加训练与部署成本。

1) Actor 网络的设计

本文所设计的 Actor 网络结构如图 9a)所示,包含 2 层隐藏层,节点数分别为 400 和 300,输出层节点数为 2。网络以无人机状态观测量 $\boldsymbol{S}$ 为输入,输出控制量 $\boldsymbol{A}$,隐藏层采用 ReLU 激活函数以便于梯度计算,输出层采用 tanh 激活函数,并通过 scaling 函数对不同控制量进行幅值缩放。



a) Actor 网络　　b) Critic 网络

图 9　Actor 与 Critic 网络结构图

2) Critic 网络的设计

Critic 网络结构如图 9b)所示,包含 3 层隐藏层,节点数分别为 400,300 和 300,输出层节点数为 1。网络以无人机状态观测量 $\boldsymbol{S}$ 和控制量 $\boldsymbol{A}$ 为输入,输出对应的 Q 值,隐藏层采用 ReLU(x) 激活函数。TD3 算法中 2 个 Critic 网络及其目标网络结构相同,但更新过程独立。

2.3.3 强侧风扰动下的轨迹控制训练

首先构建由环境和智能体构成的 DRL 训练环境,基于该环境采用 TD3 算法进行训练得到横航向轨迹控制策略。

1) 训练环境的构建

DRL 训练环境由风场、无人机系统与轨迹控制智能体构成,其详细模型见第 1 节和第 2.2 节。尽管基于高保真模型的训练具有更高可信度,但计算与时间成本较高。为提高训练效率,本文采用简化模型开展策略训练,并将所得策略迁移至高保真仿真平台进行验证。鉴于高原峡谷环境中侧风对无人机影响最为显著,本文基于横航向小扰动方程构建无人机的简化动力学模型。

$$\begin{aligned}\dot{\boldsymbol{x}}&=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{u}+\boldsymbol{F}\beta_{\mathrm{W}}\\ \boldsymbol{y}&=\boldsymbol{C}\boldsymbol{x}+\boldsymbol{G}\beta_{\mathrm{W}}\end{aligned}\tag{12}$$

式中:$\boldsymbol{x}=[\beta_{\mathrm{k}},p,r,\phi]^{\mathrm{T}}$;$\boldsymbol{u}=[\delta_{\mathrm{a}},\delta_{\mathrm{r}}]^{\mathrm{T}}$;$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别为无人机状态空间方程的系统矩阵和控制矩阵[24];β_{W} 为外界风场相对无人机的侧滑角;$\boldsymbol{F}$ 为侧风扰动矩阵,在数值上等于 $\boldsymbol{A}$ 矩阵的第一列;$\boldsymbol{y}$ 为无人机状态量观测输出,$\boldsymbol{C}$ 和 $\boldsymbol{G}$ 根据实际物理意义决定。

$\boldsymbol{F}$ 的求解公式为

$$\boldsymbol{F}=\boldsymbol{A}(:,1)\tag{13}$$

β_{W} 的近似求解方法为

$$\beta_{\mathrm{W}} = -\tan^{-1}\left(\frac{V^{\mathrm{b}}_{\mathrm{W},y}}{V_{\infty}}\right) \tag{14}$$

式中:$V^{\mathrm{b}}_{\mathrm{W},y}$ 为风速在体轴系 y 轴分量;V_{∞} 为无人机在未受扰情况下的基准真空速。(12) 式基于小扰动假设推导得到,但在极端风扰动下,部分参数受到大幅度扰动,小扰动假设失效。因此第 i 个时间节点输出变量 $\beta_{\mathrm{k},i}$ 时需基于上一时间节点输出变量 $\beta_{\mathrm{k},i-1}$ 进行修正

$$\beta_{\mathrm{k},i} = \beta_{\mathrm{k}}/\cos\beta_{\mathrm{k},i-1} \tag{15}$$

无人机在高原峡谷风场中水平轨迹受侧风影响显著,因此使用侧风模型构建飞行环境。鉴于无人机穿越峡谷时侧风通常呈现“由小到大再减小”的变化规律,本文采用图 10 的 1−cos 突风模型描述侧风,在简化建模同时保留了峡谷风场典型特征。其中:V_{WN},V_{WE},V_{WD} 分别表示风速在北向、东向和垂直方向(下方)分量。

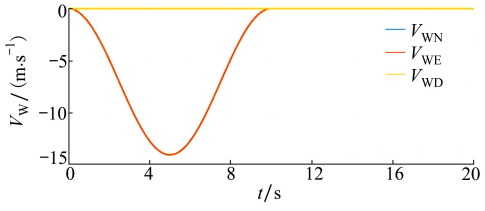


图 10　简化侧向风场模型

2) TD3 算法超参数的选择

本文采用的 TD3 算法超参数取值如表 4 所示。

表 4　TD3 算法主要超参数的取值

参数	值
Critic 网络 1 学习率	1×10^{-4}
Critic 网络 2 学习率	5×10^{-5}
Actor 网络学习率	1×10^{-5}
Actor 延迟更新步数	2
Actor 噪声方差	0.1
Actor 噪声衰减率	0.05
奖励系数	1
Q 值折扣因子	0.99

3) 训练结果分析

在上述侧风环境中对侧向轨迹控制方法进行训练,经过 1 000 幕训练后得到本文的 DRL 侧向轨迹控制策略,训练过程中单幕、5 幕平均评价值变化曲线如图 11 所示。该控制策略在第 60 幕左右就找到了高回报值区间,并在第 600 幕左右稳定到了最优值 7 附近,体现出较高的学习效率。

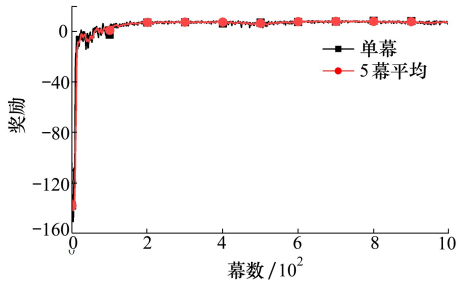


图 11　DRL 侧向轨迹控制策略训练结果

3　飞行仿真验证

3.1　DRL 与 L1 方法的数字飞行仿真对比

本节针对基于 L1 制导律的补偿式 DRL 方法,开展数字飞行仿真验证。

3.1.1　极端突风风场中的飞行仿真

假设无人机在 4 000 m 高度,以 41 m/s 的真空速定常直线平飞,在 $t=0$ 时刻遭遇图 10 所示的突风,分别采用基准的 L1 制导律和本文 DRL 控制策略,基于线性动力学模型的仿真结果如图 12 所示。

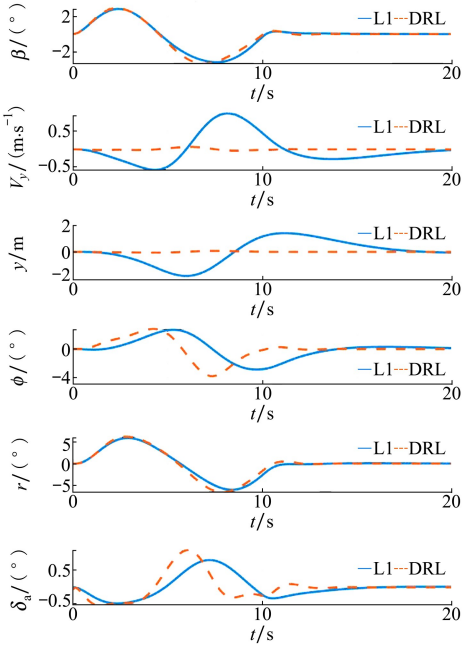


图 12　极端突风中的线性模型数字飞行仿真

由图 12 可知,基于 DRL 补偿的控制策略展现

出远超 L1 方法的优越性能。L1 方法中侧向位移偏差的最大值为 1.35 m,而 DRL 方法仅为 0.075 m。这是因为 DRL 方法具有更快的副翼响应、更积极的滚转角响应,有效抑制了偏差的初始积累。

当无人机受到侧风扰动时,图 12 中的 β 最先发生改变,进而所受力和力矩发生改变、加速度和角加速度变化,进一步改变速度、角速度,最后姿态角和位移改变。

L1 制导律基于位移和航向角偏差计算目标滚转角,其对风场扰动的响应存在一定滞后,难以及时抑制扰动初期变化。相比之下,DRL 控制策略引入侧滑角、角速度等对风场更敏感的多维状态观测量,使其能够提前感知扰动并生成具有前馈特性的补偿指令,从而实现更快速、有效的扰动抑制。

3.1.2 峡谷风场中的 6DoF 数字飞行仿真

为验证本文提出的补偿式 DRL 控制策略可行性,将其移植到 6DoF 模型中作为横航向轨迹控制器并开展飞行仿真。在第 1 节生成的风场基础上加入基于 Dryden 模型,紊流等级为“较严重的”的紊流风以模拟实际风场。仿真过程中无人机在峡谷口外垂直于峡谷飞行,感受到的风如图 13 所示。

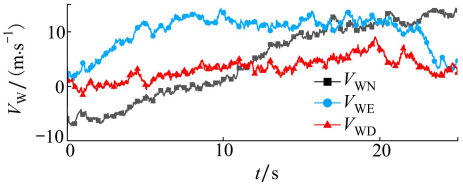


图 13 无人机感受到的风速在地轴系的分量

由图 13 可见,相较于简化线性模型,6DoF 仿真中的风场环境在 3 个方向上风速变化均更为剧烈,更加符合极端风场特征。

采用本文的补偿式 DRL 方法作为横航向轨迹控制器,仿真所得空速、侧向位移、侧向速度、副翼、油门变化曲线如图 14 所示。由图 14 可知,DRL 控制策略下的无人机侧向位移偏差不足 0.35 m,副翼的总用舵量在 8°以内,体现出该控制方法对高原峡谷风场扰动具有较好的抑制效果。此外,无人机的空速从巡航的 41 m/s 下降到最小约 30 m/s,这是因为最大的风速已经达到巡航速度的 35%。而由图 14 中 δ_t 曲线可见,油门在 8 s 左右已达到最大值,说明遭遇的风是很极端的,不过从表 2 可知失速速度为25 m/s,该控制方法仍能确保足够的安全余量。

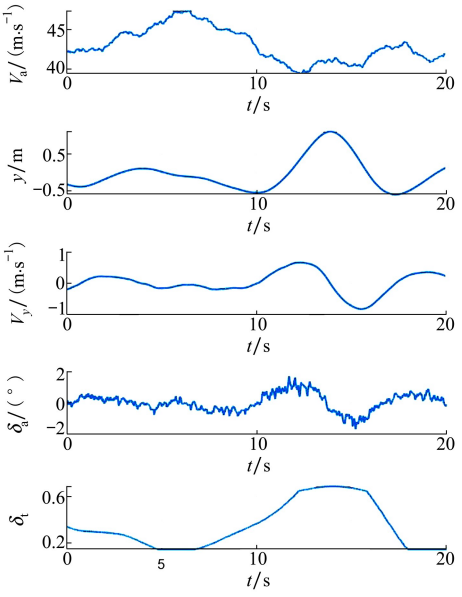


图 14 峡谷风场数字飞行仿真

基于 6DoF 非线性动力学模型的数字飞行仿真证明,基于简单风场环境训练得到的 DRL 控制方法在极端复杂风场环境下仍能展现出较好的抗风效果,体现出该控制方法具有较好的鲁棒性。

3.2 基于 DRL 抗风飞行的半物理仿真

为进一步验证控制系统物理特性对所设计的控制律的敏感程度并将其应用到真实中小型飞行器中,本文搭建了半物理仿真平台。

3.2.1 半物理仿真系统介绍

1) 仿真平台系统架构

半物理仿真系统如图 15 所示,主要包含三部分:①运行在实时仿真机中,在 Simulink 仿真平台上搭建的无人机动力学数字模型;②飞行控制器硬件在环的实时仿真平台,根据无人机模型传入的传感器信息,得到控制指令,并发送到动力学数字模型;③实景显示软件与无人机地面站,用于地面显示无

图 15 半物理仿真系统架构

人机状态、发送飞行指令等。

2) 半物理仿真的初始化

开展半物理仿真试验前,需对其进行初始化。为了对比智能化抗风控制方法的有效性,飞行控制系统预先写入飞控程序,控制架构如图 5 所示。其中,横航向轨迹控制分别写入基准的 L1 方法和本文提出的 DRL 控制方法,二者可在飞行中进行切换。半物理仿真环境中采用如图 3 所示的峡谷风场,同时叠加基于 Dryden 模型的紊流风以模拟实际风场。

3.2.2　半物理仿真结果分析

无人机以等速、定高、直线方式穿越高原峡谷风场,分别基于补偿式 DRL 方法与基准的 L1 方法开展半物理仿真试验,得到风速在地轴系分量、侧向位移、侧向速度、偏航角速度及副翼变化曲线如图 16 所示。

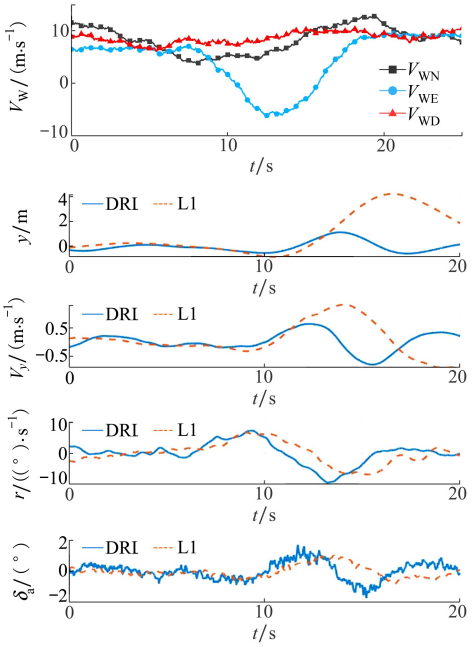


图 16　峡谷风场中的半物理仿真结果

由图 16 可知,DRL 控制方法下,无人机侧向位移扰动量在 1.2 m 以内,舵偏量在 5°以内。与图 14 中数字仿真的轨迹扰动量(y 最大为 0.3 m)相比,半物理仿真试验中扰动量显著增大,这是因为半物理仿真考虑了更真实的传感器误差、舵面的频率限制等因素,同时半物理仿真中风场扰动量更大。

仿真结果表明,在相同的风场条件下,DRL 控制策略与 L1 制导法的侧向轨迹控制性能呈现显著差异:图 16 中 DRL 方法的侧向位移最大扰动量仅为 1.2 m,相较 L1 方法(4.2 m)减小了 71.4%。

由图 16 可知,2 种方法的舵面偏转幅度相近,但 DRL 方法响应更快,能够更好地抑制初始扰动积累。这是由于 L1 控制律控制量 ϕ_c 更新依赖于滞后于风场变化的状态反馈(y,χ)。在交变风场中,该机制导致显著的相位滞后现象:当侧风强度达到峰值时,控制量未能及时响应;待风速减弱或反向时,控制量却达到极值,从而导致无人机轨迹偏差的累积。

相较之下,DRL 控制器通过融合多维度状态信息,构建更具前瞻性的扰动感知机制,有效抑制交变风场对无人机轨迹的影响。DRL 方法用舵量上与 L1 方法相当,但其敏锐的侧风变化跟踪能力显著提升了控制效果。半物理仿真结果进一步证明,该方法在复杂风场条件下不仅具备更高的控制精度,还对传感器噪声表现出良好的鲁棒性。

4　结　论

本文以 L1 制导律为基础,设计了一种面向高原峡谷风场环境的补偿式 DRL 横航向轨迹控制策略。首先,基于典型高原峡谷地形构建风场模型,并以无人机横航向线性动力学为基础,构建简化的侧风扰动环境,开展以抑制侧向偏移为目标的 DRL 策略训练。通过以 L1 制导律为基础训练策略,实现更快的收敛与更优的初始性能。同时,本文在状态观测量、评价函数等方面进行了针对性优化,加快了训练收敛速度,并进一步将所获策略迁移至 6DoF 高保真模型及半物理仿真试验平台进行验证。

结果表明,该补偿式 DRL 轨迹控制策略在高原峡谷风场扰动下表现出优良鲁棒性与控制精度。在最大侧向风速为 16 m/s 的扰动环境中,该策略控制下的最大侧向位移仅为传统 L1 方法的 28.6%。此外,本文提出方法能够以较小的训练代价得到适用于高保真模型的控制方法,展现出良好的泛化能力、迁移特性及工程应用潜力,为无人机在复杂风场环境下的稳定飞行提供一种高效可行的解决方案。

后续将考虑构建更丰富复杂的风场模型,引入风场感知预测机制,增强策略在未知环境下的适应性。此外还将考虑引入直接力控制,对横向力或加速度进行调控,以提升系统响应速度和抗扰能力。

参考文献:

[1] Zhang Q, Zhang J J, Wang X Y, et al. Wind field disturbance analysis and flight control system design for a novel tilt-rotor UAV[J]. IEEE Access, 2020, 8: 211401-211410.

[2] Wen J Y, Wang H L, Li D W, et al. Anti-wind trajectory control for unmanned aerial vehicle with nonlinear dynamic inversion based on high order sliding mode[C]//3rd International Conference on Unmanned Systems, Harbin, 2020: 1060-1065.

[3] Xing Z W, Zhang Y M, Su C Y. Active wind rejection control for a quadrotor UAV against unknown winds[J]. IEEE Trans on Aerospace and Electronic Systems, 2023, 59(6): 8956-8968.

[4] Oconnell M, Shi G, Shi X, et al. Neural-fly enables rapid learning for agile flight in strong winds[J]. Science Robotics, 2022, 7(66): 6597.

[5] Hwanggao J, Sa I, Siegwart R, et al. Control of a quadrotor with reinforcement learning[J]. IEEE Robotics and Automation Letters, 2017, 2(4): 2096-2103.

[6] Song F L, Li Z, Yang S C, et al. Anti-disturbance compensation for quadrotor close crossing flight based on deep reinforcement learning[J]. IEEE Trans on Industrial Electronics, 2023, 70(3): 3013-3023.

[7] Xue J, Liu Z, Liu G, et al. Robust wind-resistant hovering control of quadrotor UAVs using deep reinforcement learning[J/OL](2023-10-16)[2025-03-01]. https://doi.org/10.1109/TIV.2023.3324687.

[8] Song Y L, Romero A, Müller M, et al. Reaching the limit in autonomous racing: optimal control versus reinforcement learning[J]. Science Robotics, 2023, 8(82): 1462.

[9] 周晓雨, 黄江涛, 章胜, 等. 考虑强风干扰的固定翼飞行器“神经元”飞行气动建模[J]. 空气动力学学报, 2024, 42(3): 92-101.
Zhou Xiaoyu, Huang Jiangtao, Zhang Sheng, et al. Aerodynamic modeling of "neural"-fly for fixed-wing aircraft considering strong wind interference[J]. Acta Aerodynamica Sinica, 2024, 42(3): 92-101. (in Chinese)

[10] Razzaghi P, Tabrizian A, Guo W, et al. A survey on reinforcement learning in aviation applications[J]. Engineering Applications of Artificial Intelligence, 2024, 136: 108911.

[11] Rennie G. Autonomous control of simulated fixed wing aircraft using deep reinforcement learning[D]. Bath: Universityy of Bath, 2018.

[12] Kong S, Li M, Zhou Y, et al. Effective control of unmanned aerial vehicles based on reward shaping[C]//2023 6th International Conference on Robotics, Control and Automation Engineering, Suzhou, China, 2023: 135-139.

[13] Bohn E, Coates E M, Moe S, et al. Deep reinforcement learning attitude control of fixed-wing UAVs using proximal policy optimization[C]//International Conference on Unmanned Aircraft Systems, Atlanta, 2019: 523-533.

[14] Bohn E, Coates E M, Reinhardt D, et al. Data-efficient deep reinforcement learning for attitude control of fixed-wing UAVs: field experiments[J]. IEEE Trans on Neural Networks and Learning Systems, 2024, 35(3): 3168-3180.

[15] Bohn E. Reinforcement learning for optimization of nonlinear and predictive control[D]. Trondheim, Norway: Norwegian University of Science and Technology, 2022.

[16] Chowdhury M, Keshmiri S. Design and flight test validation of an ai-based longitudinal flight controller for fixed-wing UASs[C]//2022 IEEE Aerospace Conference, Big Sky, Montana, 2022: 1-12.

[17] Chowdhury M, Keshmiri S. Interchangeable reinforcement-learning flight controller for fixed-wing UASs[J]. IEEE Trans on Aerospace and Electronic Systems, 2024, 60(2): 2305-2318.

[18] Fletcher L J, Clarke R J, Richardson T S, et al. Reinforcement learning for a perched landing in the presence of wind[C]//AIAA Science and Technology Forum and Exposition, 2021: 1-14.

[19] 朱勇杰. 山区峡谷风场特性研究[D]. 焦作: 河南理工大学, 2022.
Zhu Yongjie. Research on wind field characteristics of mountain canyon[D]. Jiaozuo: Henan Polytechnic University, 2022. (in Chinese)

[20] Anderson J D. Fundamentals of aerodynamics[M]. New York: McGraw Hill, 2011.

[21] Faleiro L F, Lambregts A A. Analysis and tuning of a "total energy control system" control law using eigenstructure assignment[J]. Aerospace Science and Technology, 1999, 3(3): 127-140.

[22] Park S, Deyst J, How J P. A new nonlinear guidance logic for trajectory tracking[C]//AIAA Guidance, Navigation, and Con-

trol Conference, Rhode Island, 2004: 941-956.

[23] Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods[C]//The 35th International Conference on Machine Learning, Stockholm, Sweden, 2018: 1587-1596.

[24] 方振平, 陈万春, 张曙光. 航空飞行器飞行动力学[M]. 北京: 北京航空航天大学出版社, 2005.
Fang Zhenping, Chen Wanchun, Zhang Shuguang. Flight dynamics of aircraft[M]. Beijing: Beihang University Press, 2005. (in Chinese)

Wind disturbance-resilient flight control for small and medium-sized UAVs in plateau canyon environments using deep reinforcement learning

Zhu Yue, Wang Rui, Zhou Zhou

(School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Fixed-wing unmanned aerial vehicles (UAVs) featured in long endurance and extended range, demonstrate notable advantages in performing wide-area surveillance missions over complex terrains such as plateau canyons. However, the presence of strong and highly variable wind fields in such environments poses serious challenges to flight safety and trajectory stability. The present study focuses on lateral trajectory control in typical plateau canyon wind environments and proposes a compensation-based deep reinforcement learning (DRL) control strategy grounded in the L1 guidance law framework. To achieve both model fidelity and efficient training, the control policy is trained in an environment composed of a simplified dynamics model and a wind field model retaining key canyon characteristics, guided by a reward function tailored to lateral trajectory control. Then the trained policy is successfully transferred to a six-degree-of-freedom high-fidelity model and a hardware-in-the-loop (HIL) simulation platform for validation. The results show that the present control strategy effectively suppresses wind-induced disturbances in plateau canyon environments. Under extreme lateral wind conditions with a maximum crosswind speed of 16 m/s, the trajectory deviation is reduced to only 28.6% comparing with that by using the traditional L1 method. The results further highlight the method's strong transferability, robustness, and practical feasibility for engineering applications.

Keywords: deep reinforcement learning; TD3 algorithm; trajectory control; extreme wind fields; fixed-wing unmanned aerial vehicle (UAV)

引用格式: *朱越, 王睿, 周洲. 基于深度强化学习的中小型无人机高原峡谷抗风飞行控制[J]. 西北工业大学学报, 2026, 44(1): 01-11.*

Zhu Yue, Wang Rui, Zhou Zhou. Wind disturbance-resilient flight control for small and medium-sized UAVs in plateau canyon environments using deep reinforcement learning[J]. Journal of Northwestern Polytechnical University, 2026, 44(1): 01-11. (in Chinese)