

# 基于深度强化学习技术的舰载无人机 自主着舰控制研究

黄江涛, 刘 刚, 周 攀, 章 胜, 杜 昕

(中国空气动力研究与发展中心空天技术研究所, 四川 绵阳 621000)

[摘要] 自主着舰是未来舰载无人机面临的重要难题与关键技术。基于 TD3 算法结合舰载飞机六自由度运动以及航空母舰运动模型, 构建了交互式深度强化学习仿真环境。针对典型海况进行了舰载无人机自主着舰训练, 仿真训练过程中综合考虑海况以及航空母舰纵荡、横荡和沉浮 3 个线扰动, 滚转、俯仰和偏航 3 个角扰动等因素, 建立对应简化运动模型; 基于某型飞机气动数据进行气动力建模, 建立六自由度运动学/动力学模型; 基于 TD3 强化学习算法, 结合前馈型深度神经网络技术, 在高性能 GPU 工作站上建立舰载机着舰交互训练环境。通过某型舰载无人机在无模型环境中“试错”训练, 验证了 AI 技术在舰载无人机自主着舰控制中的可行性。

[关键词] 强化学习, 舰载无人机, 智能着舰, 舵偏指令, 深度神经网络

[中图分类号] V211.3 [文献标志码] A [文章编号] 1672-1292(2022)03-0063-09

## Research on Autonomous Landing Control of Carrier-borne UAV Based on Deep Reinforcement Learning Technology

Huang Jiangtao, Liu Gang, Zhou Pan, Zhang Sheng, Du Xin

(Aerospace Technology Institute, China Aerodynamics Research and Development Center, Mianyang 621000, China)

**Abstract:** Autonomous landing is an important problem and a key technology for future Carrier-borne UAV. Based on the TD3 algorithm, combined with the 6-DOF motion model of carrier aircraft and the motion model of aircraft carrier, an interactive deep reinforcement learning simulation environment is constructed. In the process of simulation training, the corresponding simplified motion model is established by considering the sea conditions, three line disturbances of aircraft carrier including surge, sway and heave, and three angular disturbances of roll, pitch and yaw. Based on the aerodynamic data of a certain type of aircraft, the aerodynamic model is established, and the six degree of freedom dynamics model is also established. Based on TD3 reinforcement learning algorithm, this paper further introduces an auxiliary network, an adaptive variance and learning step adjustment algorithm to accelerate convergence and improve training stability. Furthermore, combined with feed forward deep neural network technology, an interactive training environment for carrier based aircraft landing is established on high performance GPU workstation. Through the "trial and error" training of a certain type of Carrier-borne UAV in interactive environment, the feasibility of AI technology in Carrier-borne UAV autonomous landing control is verified.

**Key words:** reinforcement learning, carrier-borne UAV, intelligent carrier landing, control surface command, deep neural network

舰载机着舰时由于大气湍流、舰尾流、甲板运动和海浪等随机复杂环境因素<sup>[1-3]</sup>, 对着舰精度及成功率产生了严重影响, 着舰难度极大, 因此, 舰载机着舰过程被称为“刀尖上的舞蹈”。舰载无人机的着舰难度更大, 目前大多数舰载无人机是通过遥控方式实现整个过程的控制<sup>[4-5]</sup>。随着技术的不断进步, 各海军强国要求舰载无人机能够实现航母上的自主起降。因此, 对于先进自主着舰技术的研究显得十分重要。

随着人工智能技术的迅速发展, 无人作战飞行器将具备自主决策能力, 其决策反应时间远超人类飞行

员. 基于“试错”训练学习的 AI 技术能够实现从状态观测到舵面/推力控制的一体化飞行智能控制决策, 可有效消除传统控制技术航迹/姿态控制双回路带来的性能损失, 实现“制导/控制高度一体化”, 有望为舰载无人机自主着舰提供更为有效的智能决策技术支撑.

目前, 基于深度强化学习的无人机控制大多集中在 PID 参数训练上<sup>[6]</sup>, 无模型交互以及基于 AI 技术的舰载无人机自主着舰控制的研究较少, 处于刚起步阶段. 余洋<sup>[7]</sup>基于行动者-评论家模型开展了航空母舰理想静止状态以及考虑简单风速环境的自主着舰研究. 吴昭欣等<sup>[8]</sup>基于深度强化学习进行了舰载机三自由度近引导研究. 现有研究基础在着舰交互环境及动力学模型上不够完善, 研究不够系统.

本文基于某型舰载机六自由度运动方程/航母运动模型, 综合考虑风场环境, 结合深度强化学习算法构建交互训练环境, 对舰载无人机智能着舰进行了研究.

## 1 舰载无人机运动学/动力学建模

### 1.1 舰载无人机气动力建模

本文采用某型无人机气动力数据库进行建模. 由于舰载机着舰过程中力图避免剧烈机动动作, 且对迎角、侧滑角变化范围进行严格限制, 因此, 完全可以采用传统多项式方法进行无量纲气动力建模.

升力系数表达式:

$$C_L = \begin{cases} A_0 + A_1\alpha + A_2\delta_e, & \alpha \leq \alpha_L; \\ A'_0 + A'_1\alpha + A'_2\alpha^2 + A'_3\delta_e, & \alpha > \alpha_L. \end{cases} \quad (1)$$

阻力系数表达式:

$$C_D = \begin{cases} B_0 + B_1\alpha + B_2\alpha^2 + B_3\delta_e, & \alpha \leq \alpha_L; \\ B'_0 + B'_1\alpha + B'_2\alpha^2 + B'_3\delta_e, & \alpha > \alpha_L. \end{cases} \quad (2)$$

侧向力系数表达式:

$$C_Y = C_0 + C_1\beta + (C_2\alpha + C_3)\delta_a + C_4\delta_r. \quad (3)$$

滚转力矩系数表达式:

$$C_l = \begin{cases} (D_1\alpha + D_2)\beta + D_3p + (D_4\alpha + D_5)\delta_e + D_6r + (D_7\alpha + D_8)\delta_r, & \alpha \leq \alpha_L; \\ (D'_1\alpha + D'_2)\beta + D'_3p + (D'_4\alpha + D'_5)\delta_e + D'_6r + (D'_7\alpha + D'_8)\delta_r, & \alpha > \alpha_L. \end{cases} \quad (4)$$

俯仰力矩系数表达式:

$$C_m = E_0 + E_1\alpha + E_2\delta_e + E_3q. \quad (5)$$

偏航力矩系数表达式:

$$C_n = \begin{cases} F_1\beta + F_2r + (F_3\alpha + F_4)\delta_a + F_5r + (F_6\alpha + F_7)\delta_r, & \alpha \leq \alpha_L; \\ F'_1\beta + F'_2r + (F'_3\alpha + F'_4)\delta_e + F'_5r + (F'_6\alpha + F'_7)\delta_r, & \alpha > \alpha_L. \end{cases} \quad (6)$$

式中,  $\alpha, \beta, \delta_e, \delta_a, \delta_r, p, q, r, \alpha_L$  分别为迎角、侧滑角、升降舵舵偏、副翼舵偏、方向舵舵偏、三轴角速度以及临界迎角角度;  $A, B, C, D, E, F$  为多项式系数.

### 1.2 舰载无人机六自由度动力学模型

固定翼类无人飞行器的运动动力学模型包含质心动力学、绕质心转动运动学方程, 其模型分别为<sup>[9]</sup>:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \mathbf{R}_b^g \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \quad (7)$$

$$\begin{bmatrix} \dot{u} \\ \dot{v} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} vr - wq - g\mathbf{R}_g^b(1, 3) + \frac{X + T_x}{m} \\ -ur + wp - g\mathbf{R}_g^b(2, 3) + \frac{Y + T_y}{m} \\ uq - vp - g\mathbf{R}_g^b(3, 3) + \frac{Z + T_z}{m} \end{bmatrix}. \quad (8)$$

为防止欧拉角表示的绕质心转动运动学模型出现奇异,本文采用基于四元数法来描述绕质心转动运动学方程。

该型舰载机所采用的气动力模型假设发动机推力通过重心,无推矢作用。控制量包括升降舵、副翼、方向舵、发动机推力共 4 个量,控制周期为  $\Delta T$ 。

气动力建模所需的迎角、侧滑角信息可以通过体系下表示的速度  $\mathbf{v}=[u \ v \ w]^T$  求解:

$$\begin{aligned} V &= \sqrt{u^2 + v^2 + w^2}, \\ \alpha &= \arctan(w/u), \\ \beta &= \arcsin\left(\frac{v}{\sqrt{u^2 + v^2 + w^2}}\right). \end{aligned}$$

## 2 自主着舰环境建模

### 2.1 航空母舰扰动运动建模

在不同海况中,受海浪的影响,舰船将产生六自由扰动,即 3 个线位移扰动和 3 个旋转角位移扰动。本文研究中等海况条件下 AI 技术的可行性与效果,主要考虑垂荡和纵摇。由于侧向着舰精度主要受横摇和横荡运动的影响,综合考虑中等海况航母横荡幅值与引导控制系统侧向偏差限制,航母的横侧向运动并不会导致着舰的失败<sup>[3]</sup>,纵摇导致的纵向位移为小量,在上述条件下,可以将不同海况下的沉浮运动表达为:

$$\Delta Z = -\Delta Z_b + L_D \sin\theta + Y_D \sin\varphi, \quad (9)$$

式中,  $\Delta Z_b$  为航空母舰质心沉浮位移;  $L_D$  是着舰点距质心的纵向距离;  $Y_D$  是着舰点距质心的侧向距离;  $\theta$ 、 $\varphi$  分别为航空母舰纵摇、横摇角度。

### 2.2 风场扰动环境建模

风场建模综合考虑自由大气紊流分量、雄鸡尾流分量、舰尾流周期性及其随机分量。在不同海况中,针对雄鸡尾流,采用对军用规范 8785C 给出的雄鸡尾流图按  $X$  进行分段线性化处理后的数学模型<sup>[10]</sup>;针对舰尾流周期性及其随机分量,采用文献[3]提供的建模方式。

## 3 无模型深度强化学习环境

### 3.1 TD3 强化学习算法

本文采用 TD3 算法<sup>[11]</sup>进行强化学习训练,其核心基础是 DDPG 算法(deep deterministic policy gradient)<sup>[12]</sup>。

DDPG 算法采用 Actor-Critic 架构,能够解决“连续动作”类型强化学习问题,是一种确定性策略梯度算法,其中 Critic 评价、Actor 动作策略近似表达均采用前馈型深度神经网络。

TD3 算法通过引入双延迟目标网络,避免目标网络预估过高引起的强化学习剧烈振荡或发散,且 Actor 网络的更新要迟滞于 Critic 网络,进一步提高 Critic 网络的指导能力以及 Actor 网络决策的可靠性。

图 1 给出了采用本文建立的训练学习系统对典型的追击-逃逸问题的训练效果,其中红方飞机(智能体)训练过程中采用六自由度运动学/动力学方程;为简化训练计算量,蓝方飞机(目标机)采用三自由度运动方程。蓝方飞机在做机动逃逸过程中,红方智能体飞机能够迅速占据后向与高度优势,验证了训练系统的可靠性。

### 3.2 深度神经网络

深度强化学习研究中,采用前馈型深度神经网络<sup>[13]</sup>来表征连续系统下的状态空间到动作空间的函数关系,解决状态量“维度障碍”问题<sup>[14]</sup>。通过求解目标函数梯度并进行反向传播<sup>[15]</sup>来进行训练。

前馈型全连接神经网络结构示意图如图 2 所示,文中 Actor 网络采用 10 层隐层结构,分别为 256、512、1 024、1 024、512、256;Critic 网络采用相同的隐层结构。

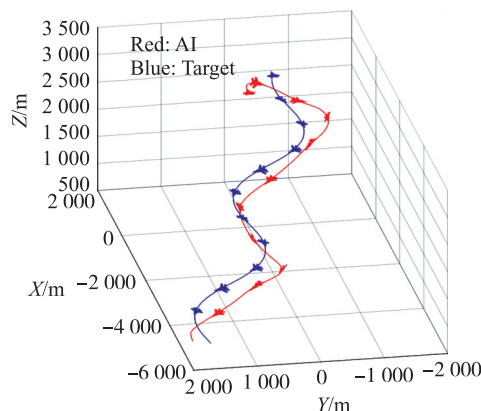


图 1 红蓝追击-逃逸仿真

Fig. 1 Pursuit escape simulation

### 3.3 马尔科夫决策过程建模

智能体与环境的交互过程通常被表达为马尔科夫决策过程(MDP)<sup>[15]</sup>. MDP 模型的构建是求解强化学习问题的重要方式. MDP 建模包含状态空间、动作空间及奖励函数等重要要素,对应智能舰载无人机自主着舰问题中的机舰关系、操纵面/发动机油门特性以及适应着舰环境的奖励设计.

对于六自由度着舰问题,状态量基本能够反映舰机相对关系以及舰载机的飞行状态. 因此,状态空间选择机舰纵向相对速度、纵向相对位置、对中线侧偏距、对中线下偏距、侧偏移速度、下偏移速度、滚转角、偏航角、俯仰角、三轴角速率、迎角、侧滑角等 14 个变量作为状态量.

舰载机动作空间包含升降舵、副翼、方向舵、发动机油门控制 4 个要素,本文训练仿真过程未考虑舵机延迟特性.

奖励函数设计综合考虑舰载机的下沉率、俯仰角、下滑道保持、迎角/侧滑角抑制等要素,构造以下奖励函数:

(1) 下沉率奖励函数主要考虑起落架承受的垂向触舰速度限制:

$$R_1 = \sqrt{\frac{|V_z - V_L|}{4|V_L|}}.$$

(2) 俯仰角奖励函数主要考虑舰载机能够以合理的俯仰姿态触舰:

$$R_2 = \sqrt{\frac{|\theta - \theta_0|}{2|\theta_0|}}.$$

(3) 下滑道保持函数主要考虑舰载机能够以正确的相对航迹着舰,这里从两方面要素设计奖励函数,即下滑线偏移距离以及相对速度矢量与相对下滑线的一致性:

$$R_3 = R_{31} + R_{32},$$

$$R_{31} = \sqrt{\frac{|Z_f - \Delta X \tan(\Gamma)|}{D_0}} + \sqrt{\frac{|y_f|}{D_0}},$$

$$R_{32} = \sqrt{\frac{|\arccos(V_{rx}R_{nx} + V_{ry}R_{ny} + V_{rz}R_{nz})|}{\vartheta_0}},$$

式中,  $Z_f$ 、 $\Delta X$ 、 $\Gamma$ 、 $y_f$ 、 $D_0$ 、 $\vartheta_0$  分别表示舰载机绝对高度、相对距离、相对下滑角、侧偏距、距离无量纲参考量、角度无量纲参考量;  $V_{rx}$ 、 $V_{ry}$ 、 $V_{rz}$  与  $R_{nx}$ 、 $R_{ny}$ 、 $R_{nz}$  分别对应舰载机相对航母的速度分量以及相对下滑线矢量.

(4) 侧滑角奖励函数主要考虑舰载机避免由于侧滑带来荷兰滚特性而导致的航迹振荡,  $\beta_0$  表示侧滑角度无量纲参考量:

$$R_4 = -\sqrt{\frac{|\beta|}{\beta_0}},$$

式中,  $\beta_0$  表示侧滑角度无量纲参考量.

(5) 其他奖励设置. 为加速训练过程,同时避免极限状态的产生,当训练过程中出现迎角/侧滑角、下滑线偏距大于指定阈值等,认为训练失败,提前结束训练同时给予以下惩罚:

$$R_5 = -\frac{|k - k_{\max}|}{100},$$

式中,  $k$ 、 $k_{\max}$  分别表示满足提前结束训练条件的仿真步数以及最大仿真步数. 该项设置能够加速实现智能体向正确方向学习效果.

综合以上奖励设置,最终采用以下奖励函数:

$$R = \frac{(R_1 + R_2 + R_3)}{4} + 0.1R_4 + R_5.$$

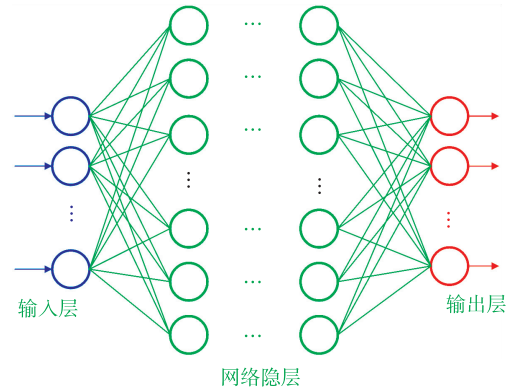


图 2 前馈型深度神经网络

Fig. 2 Feedforward deep neural network



## 4 基于 AI 技术的舰载无人机自主着舰仿真验证

对舰载机的初始位置、初始速度以及航空母舰的初始速度随机取样进行强化学习. 图 3、图 4 分别给出了强化学习历程平均回报以及终止步回报,由于引入提前结束训练惩罚,且惩罚与结束步长相关,因而在训练初期阶段,回报函数呈现不规则振荡. 图 5、图 6 所示为典型步数的决策能力测试,可以看出,在训练前期,智能体不具备正确决策能力,处于漫无目的的飞行状态,且飞行过程中常处于大迎角与侧滑角状态;经过 17 万次强化训练学习,智能体逐渐具备一定自主着舰控制能力,但控制精度较低,呈现明显的振荡现象. 经过 40 万次左右训练,智能体具备自主着舰能力.

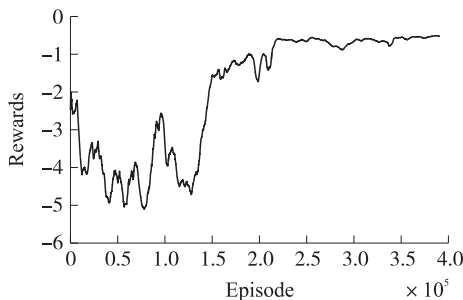


图 3 强化学习过程平均奖励

Fig. 3 Average reward in reinforcement learning process

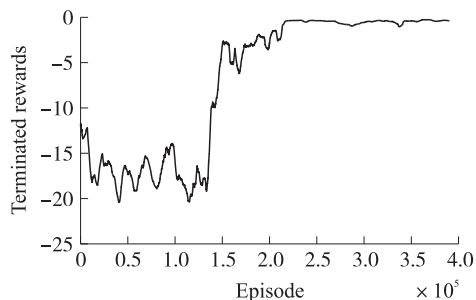


图 4 强化学习过程终止条件奖励

Fig. 4 Reward for termination condition of reinforcement learning process

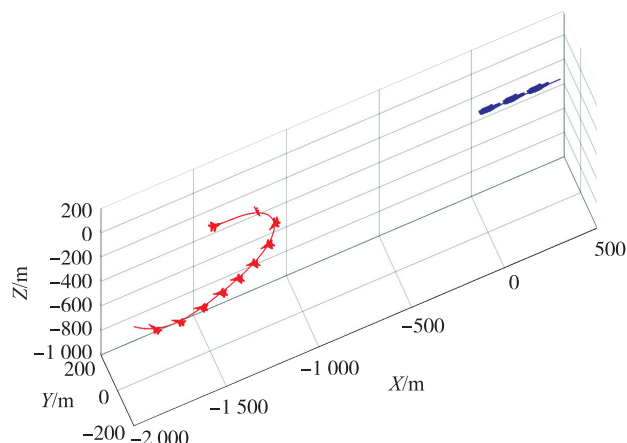


图 5 智能体训练初期仿真

Fig. 5 Simulation at the beginning of agent training

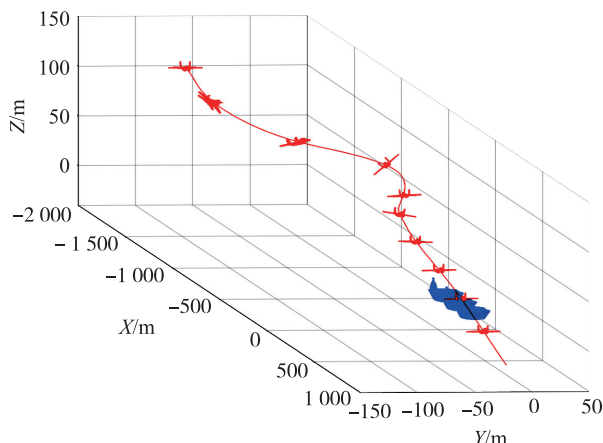


图 6 智能体 17 万次训练仿真

Fig. 6 Simulation of agent landing after 170 000 training

### 4.1 航空母舰运动条件下的着舰仿真

以航空母舰随机初始化运动速度为例,验证强化学习结果的有效性,航母的运动速度处于 15~30 m/s 之间. 图 7 给出了航母理想运动条件下的着舰随机仿真,所训练的智能体能够保持与下滑道的一致,实现成功着舰.

图 8 所示为舰载机着舰过程中迎角、侧滑角、侧偏距的变化历程. 10 s 以后舰载机能够准确对准着舰点,沿下滑道稳定飞行,迎角稳定,侧滑角接近于 0,侧偏距控制误差低于 0.15 m. 图 9 给出了 3 个姿态角的变化历程,着舰过程中处于舰载机趋于稳定的控制状态;图 10 为舵偏指令,所训练决策机的舵偏指令历程光滑合理;图 11 给出了舰载机地速变化情况,可以看到,着舰侧偏速度接近于 0,触舰垂向速度低于 7 m/s,满足着舰需求.

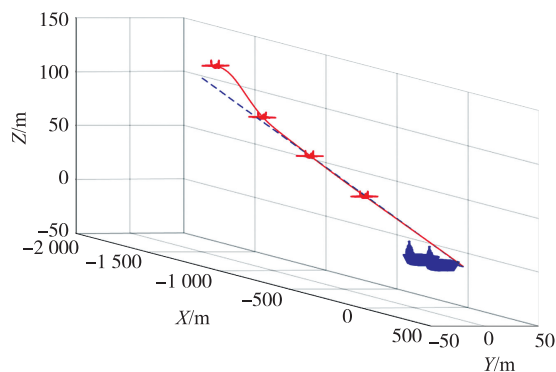


图 7 航母理想运动条件下的着舰仿真

Fig. 7 Landing simulation under ideal motion conditions

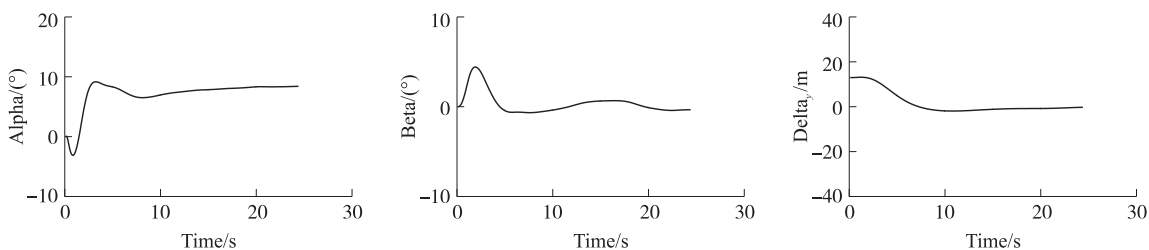


图 8 理想运动条件下的迎角/侧滑角/侧偏距历程仿真

Fig. 8 Simulation of attack angle/sideslip angle/sideslip distance under ideal motion conditions

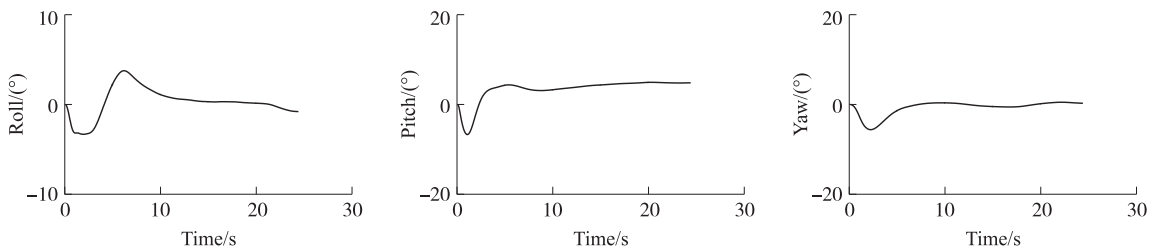


图 9 理想运动条件下的姿态角变化历程仿真

Fig. 9 Simulation of attitude angle history under ideal motion conditions

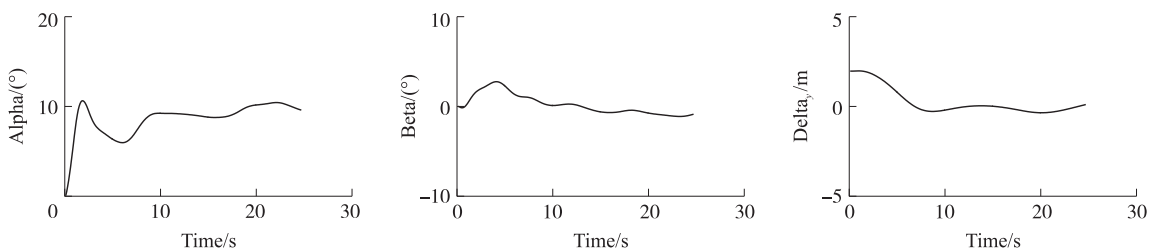


图 10 迎角/侧滑角/侧偏距历程仿真

Fig. 10 Simulation of attack/sideslip angle/sideslip distance

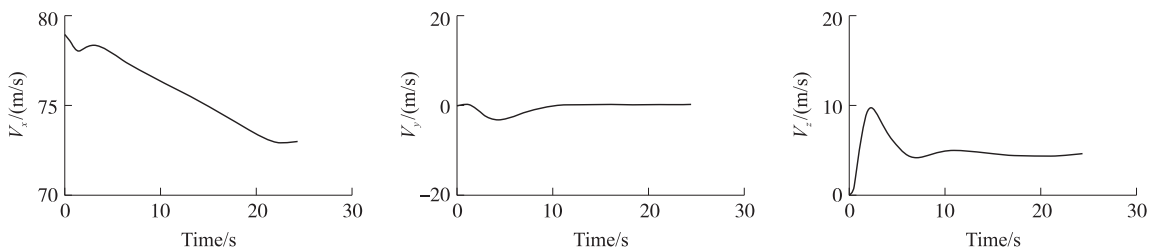


图 11 地速变化历程

Fig. 11 History of earth velocity change

#### 4.2 航空母舰扰动条件下的着舰仿真

航母的运动速度在 15~30 m/s 之间进行随机初始化. 图 12 给出了风场环境扰动条件下的仿真结果,智能体能够实现成功着舰.

图 13 给出了着舰过程中迎角、侧滑角、侧偏距的变化历程,由于奖励函数设计主要保持俯仰角度尽量控制在 5°左右,因此为实现稳定飞行,迎角呈现振荡,侧滑角小幅度变化,侧偏距误差小于 0.15 m;从图 14 姿态角变化历程可以看到,俯仰角有小幅度变化,滚转、偏航接近于 0;三通道舵偏指令有效抑制了风环境扰动对横航向的干扰,如图 15 所示;图 16、17 给出了舰载机地速变化以及对甲板跟踪状况,着舰侧偏速

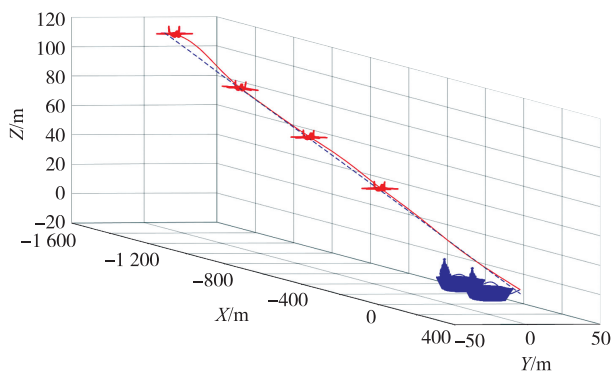


图 12 航母扰动条件下的着舰仿真

Fig. 12 Landing simulation under carrier disturbance

度振荡幅值不超过 0.15 m/s,触舰垂向速度呈现振荡现象,且低于 7 m/s;满足着舰需求。

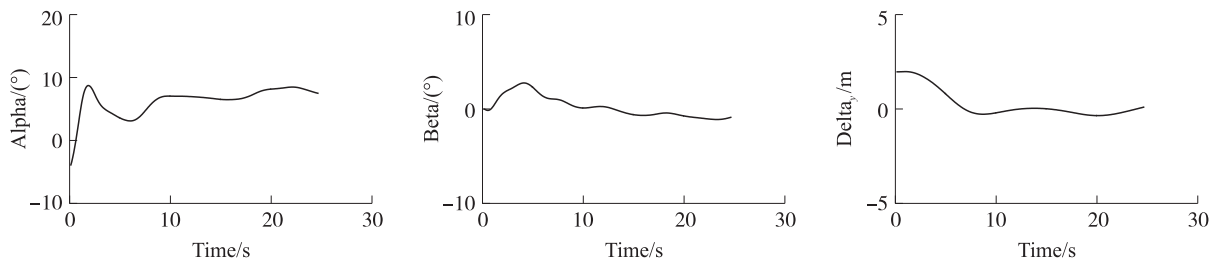


图 13 航母扰动条件下的迎角/侧滑角/侧偏距历程仿真

Fig. 13 Simulation of attack angle/sideslip angle/sideslip distance under carrier disturbance

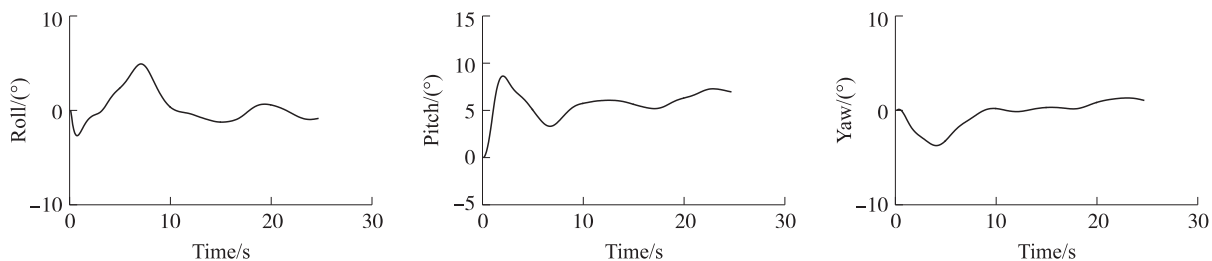


图 14 航母扰动条件下的姿态角变化历程仿真

Fig. 14 Simulation of attitude angle history under carrier disturbance

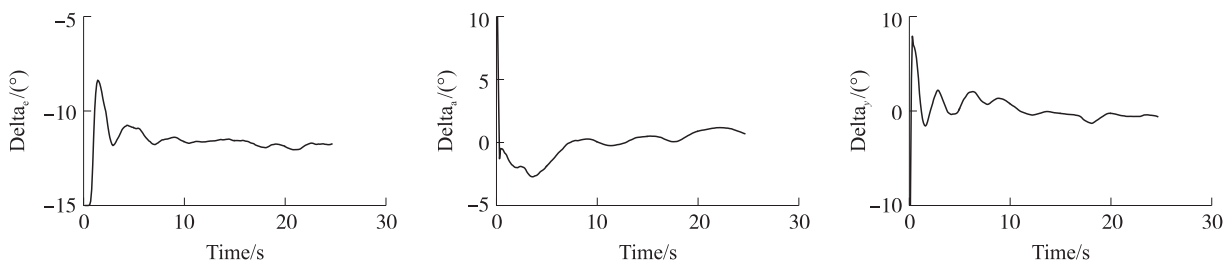


图 15 航母扰动条件下的舵偏指令

Fig. 15 Control surface command under carrier disturbance

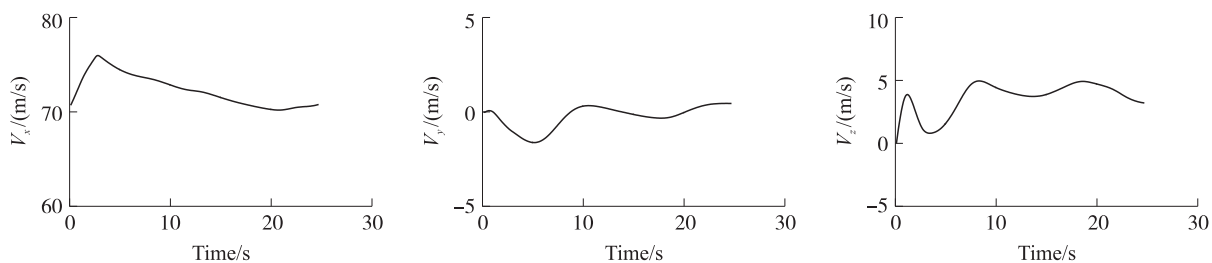


图 16 航母扰动条件下的地速变化历程

Fig. 16 Ground speed change history under carrier disturbance

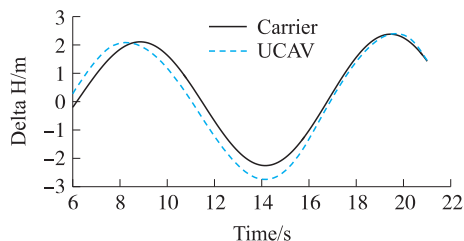


图 17 航母扰动条件下智能体对甲板运动的跟踪

Fig. 17 Tracking of deck motion by agents under carrier disturbance

### 4.3 风场环境扰动条件下的着舰仿真

风场扰动主要考虑了雄鸡尾流以及纵摇诱导尾流的影响,由于随机紊流对系统的影响远远小于前者,故本文暂不考虑随机紊流,这并不影响考核神经网络决策机的有效性.图 18 给出了风场环境扰动条件下着舰仿真,在风场环境扰动条件下,智能体能够实现成功着舰.

图 19 给出了着舰过程中迎角、侧滑角、侧偏距的变化历程,由于奖励函数设计主要保持俯仰角度尽量控制在  $5^\circ$  左右,因此为实现稳定飞行,迎角呈现振荡,侧滑角小幅度变化,侧偏距误差小于  $0.15\text{ m}$ ;从图 20 给出的姿态角变化历程可以看到,在触舰前俯仰角有小幅度变化,滚转、偏航接近于 0,这是由于三通道舵偏指令有效抑制了风环境扰动对横航向的干扰,如图 21 所示;图 22 给出了舰载机地速变化,着舰侧偏速度振荡幅值不超过  $0.15\text{ m/s}$ ,着舰垂向速度呈现振荡现象,且低于  $7\text{ m/s}$ ;满足着舰需求.

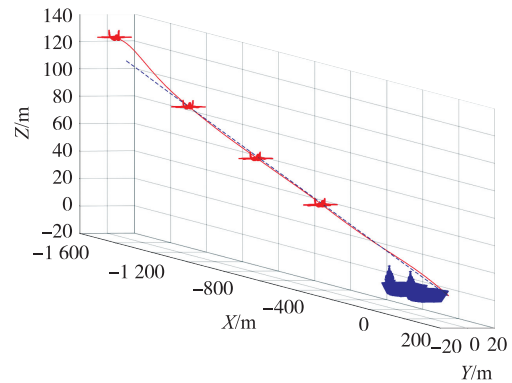


图 18 风场环境扰动条件下的着舰仿真

Fig. 18 Landing simulation under wind disturbance

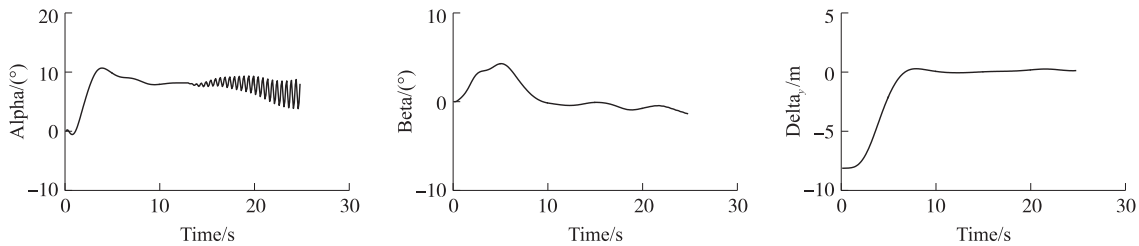


图 19 风场环境扰动条件下的迎角/侧滑角/侧偏距历程仿真

Fig. 19 Simulation of attack angle/sideslip angle/sideslip distance under wind disturbance

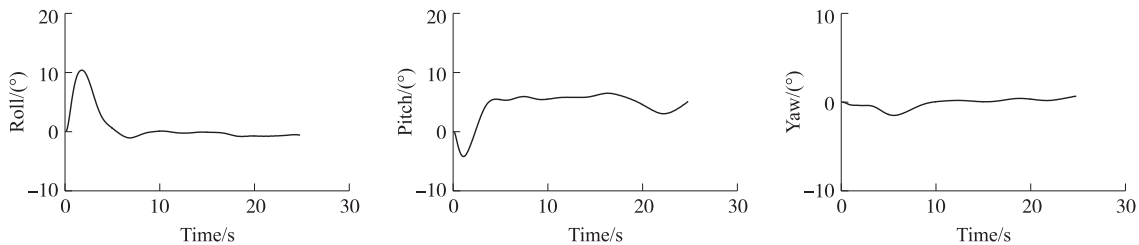


图 20 风场环境扰动条件下的姿态角变化历程仿真

Fig. 20 Simulation of attitude angle history under wind disturbance

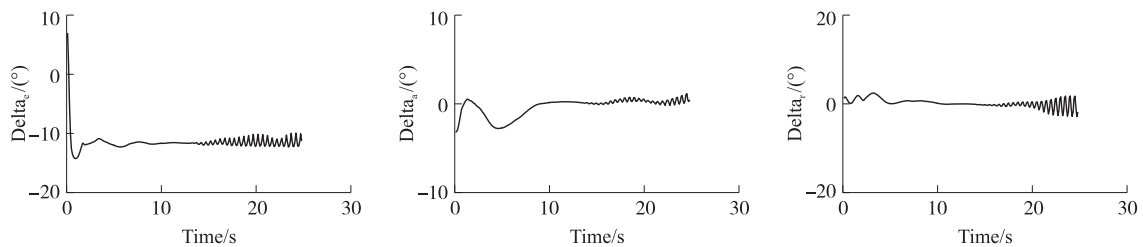


图 21 风场环境扰动条件下的舵偏指令

Fig. 21 Control surface command under wind disturbance

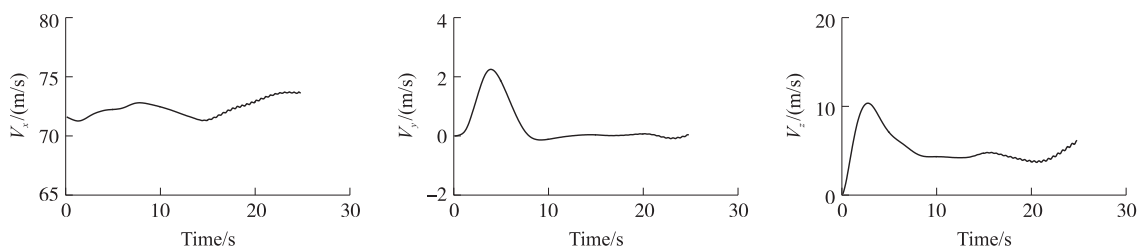


图 22 风场环境扰动条件下的地速变化历程

Fig. 22 Ground speed change history under wind disturbance



## 5 结论

本文针对不同着舰条件,基于深度强化学习开展了舰载无人机自主着舰控制研究,初步验证了 AI 技术在自主着舰控制中的可行性:

(1) 建立的智能体自主着舰交互强化学习系统具有较好的稳定性;

(2) 所设计的观测量能够为神经网络决策机输出提供较为完备的状态信息;所设置的奖励函数可为智能体与着舰环境交互提供有效指导,能够充分评价机舰关系以及舰载机姿态;

(3) 智能体能够实现自主着舰,完全胜任不同起始位置、相对速度等条件下的着舰任务,着舰侧偏距误差满足着舰要求,具有较好的适应性;

(4) 在分别考虑航空母舰扰动、风场环境下,检验了智能体在中等海况下自主着舰的有效性,决策机舵偏指令光滑合理。

本文研究主要针对智能体对航母运动、甲板扰动、风场环境等海况因素的适应性进行独立验证,在真实环境中,需考虑舵机延迟特性,且几个要素相互耦合影响,因此,随机因素及状态量选取将更为复杂,对强化学习的计算量、稳定性提出了更高要求。因此,下一步将开展更加逼近真实环境的舰载无人机强化学习算法改进、适应性验证等研究工作。

## [参考文献] (References)

- [1] 安军. 航母尾流模拟及舰载机着舰控制的初步研究[D]. 武汉:华中科技大学,2012.
- [2] 巩鹏潇,詹浩,柳子栋. 舰尾流影响下的舰载机着舰控制与仿真研究[J]. 航空工程进展,2013,4(3):339-345,357.
- [3] 张孝伟. 飞翼舰载无人机着舰控制技术[D]. 南京:南京航空航天大学,2017.
- [4] 郑峰婴,龚华军,甄子洋. 基于坐标系动态变化的无人机着舰引导算法[J]. 中南大学学报(自然科学版),2016,47(8):2685-2693.
- [5] ANDERSON M R. Inner and outer loop manual control of automatic carrier landing[C]//Proceedings of 1996 Guidance, Navigation, and Control Conference(AIAA). San Diego, USA: AIAA, 1996.
- [6] 黄旭,柳嘉润,贾晨辉,等. 深度确定性策略梯度算法用于无人飞行器控制[J]. 航空学报,2021,42(11):397-407.
- [7] 余洋. 面向飞行器自主着舰问题的行动者-评论家算法模型研究与实现[D]. 北京:北京交通大学,2019.
- [8] 吴昭欣,李辉,王壮,等. 基于深度强化学习的智能仿真平台设计[J]. 战术导弹技术,2020(4):193-200.
- [9] 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.
- [10] GARNETT T S. Investigation to study the aerodynamic ship wake turbulence generated by a DD963 Destroyer:ADA083663[R]. Philadelphia, USA: Boeing Vetrol Co., 1979.
- [11] COLAS C, SIGAUD O, OUDEYER P Y. GEP-PG: decoupling exploration and exploitation in deep reinforcement learning algorithms[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018.
- [12] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. Computer Science, 2015, 8(6): A187.
- [13] 焦李成,赵进,杨淑媛,等. 深度学习、优化与识别[M]. 北京:清华大学出版社,2017.
- [14] 谭浪. 强化学习在多智能体对抗中的应用研究[D]. 北京:中国运载火箭技术研究院,2019.
- [15] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. Cambridge, England: MIT Press, 2018.

[责任编辑:严海琳]