http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0403

基于强化学习的舰载机着舰直接升力控制技术

柳仁地，江驹*，张哲，刘翔
(南京航空航天大学 自动化学院，南京 211106)

摘　　要：针对舰载机自动着舰过程中受甲板运动及舰尾流扰动很容易发生触舰危险的问题，提出了基于近端策略优化(PPO)算法的舰载机自动着舰直接升力控制方法。PPO控制器以俯仰角、高度、航迹倾斜角、俯仰角速率、高度误差和航迹倾斜角速率等6个状态变作为输入，以襟翼的舵偏角增量作为输出，实现舰载机在着舰时航迹倾斜角的快速响应。与传统控制器相比，PPO控制器中的Actor-Critic框架大大提高了控制量的计算效率，降低了参数优化的难度。仿真实验基于MATLAB/Simulink中的F/A-18飞机动力学/运动学模型。利用PyCharm平台上构建的深度强化学习训练环境，通过UDP通信实现2个平台之间的数据交互。仿真结果表明：所提方法具有响应速度快、动态误差小的特点，能够将着舰的高度误差稳定在±0.2 m以内，具有较高的控制精度。

关 键 词：舰载机自动着舰；深度强化学习；近端策略优化算法；直接升力控制；UDP通信
中图分类号：V249.1
文献标志码：A　　**文章编号**：1001-5965(2025)06-2165-11

舰载机着舰是一项非常复杂的任务，由于着舰过程中受大气扰动、舰尾流、甲板运动、波浪和海况等环境因素的影响，舰载机的着舰精度与成功率受到很大影响[1]。

自从直接升力控制技术被应用在着舰过程中[2]，着舰的安全性和稳定性得到了提高。目前直接升力控制主要以PID控制、模糊控制和自抗扰等方法为主[3]。文献[4]通过基于直接升力的非线性综合解耦控制器进行着舰控制。文献[5]提出了一种动态分配控制方法，使襟翼像其他舵面一样能够快速驱动。文献[6]提出了基于移动路径跟踪(moving path following, MPF)法则的着舰引导律和一种新的集成式直接升力控制方法。文献[7]利用动态逆反馈控制方法设计飞行控制器，通过控制分配技术实现特定飞行操纵。虽然上述方法具有良好的控制效果，但对于六自由度飞机模型，都需要进行精确的数学建模和参数设计，特别是当被控对象处于强耦合、强非线性和强时变环境时，参数的调节和优化非常困难，控制效果受到很大影响。

随着AlphaGo的问世，人工智能技术得到了迅速发展，其中以强化学习为代表的算法雨后春笋般涌现[8]。再加上智能计算、并行计算等计算能力的高效提升，基于智能体与环境交互的强化学习算法在任务分配、路径规划、机动决策[9]、协同对抗、飞行控制等领域得到了广泛的应用。文献[10]采用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法实现了固定翼飞机在平飞状态下的俯仰姿态控制。文献[11]设计了基于DDPG算法的控制器实现航天器的姿态控制。文献[12]在预设性能控制器的基础上引入强化学习算法，用于评估系统性能。文献[13]利用建立好的专家经验数据集，通过模仿强化学习设计了固定翼飞机的姿态控制器。文献[14]提出了一种基于TD3算法的无人机近距空战格斗自主决策模型。文献[15]针对舰载无人机，通过TD3算法设计了全自动着舰控制器。

收稿日期：2023-06-21；**录用日期**：2023-11-10；**网络出版时间**：2023-11-24 17:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20231123.1537.005
*通信作者. E-mail：jiangju@nuaa.edu.cn

引用格式：*柳仁地，江驹，张哲，等. 基于强化学习的舰载机着舰直接升力控制技术[J]. 北京航空航天大学学报，2025，51(6)：2165-2175. LIU R D，JIANG J，ZHANG Z，et al. Direct lift control technology of carrier aircraft landing based on reinforcement learning[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2165-2175 (in Chinese).*

基于上述分析，本文将深度强化学习应用于直接升力控制中，提出一种基于近端策略优化(proximal policy optimization, PPO)算法[16-17]的自动着舰纵向控制器设计方法。在构建完Actor-Critic框架后，设计适当的奖励函数引导舰载机学习襟翼控制策略。仿真结果表明，训练后的纵向自动着舰控制器能够根据不同状态输出相应的襟翼偏转指令，实现飞行轨迹的快速变化，从而使舰载机安全稳定地着舰。

1 舰载机、甲板运动及舰尾流建模

1.1 舰载机模型建立

F/A-18舰载机的部分气动参数以及外形参数如表1所示。

表1 气动参数及外形参数

Table 1 Pneumatic parameters and shape parameters

参数	数值
机翼面积 S_w/m^2	37.16
平均气动弦长 $\bar{c}$/m	3.51
翼展 b/m	11.43
飞机质量 m/kg	16 651
绕O_bx_b轴转动惯量 I_{xx}/(kg·m^2)	31 184
绕O_by_b轴转动惯量 I_{yy}/(kg·m^2)	205 130
绕O_bz_b轴转动惯量 I_{zz}/(kg·m^2)	230 415
绕O_bx_b轴惯性积I_{xz}/(kg·m^2)	−4 028

舰载机在着舰过程中，主要受重力、气动力以及发动机推力的作用。假设发动机的安装角为0°，则推力公式为

$$T=\delta_T T_{\max} \tag{1}$$

式中：δ_T为油门开度；$T_{\max}=f(h,Ma)$为最大推力，与舰载机当前的高度和空速有关。

气动力在气流坐标系中可以分解为阻力D、侧力Y、升力L，气动力矩在机体坐标系中可以分解为滚转力矩$\bar{L}$、俯仰力矩M、偏航力矩N，气动力和气动力矩的表达式如下：

$$\begin{cases} L=\dfrac{1}{2}\rho V^2 S_w C_L \\ D=\dfrac{1}{2}\rho V^2 S_w C_D \\ Y=\dfrac{1}{2}\rho V^2 S_w C_Y \end{cases} \tag{2}$$

$$\begin{cases} \bar{L}=\dfrac{1}{2}\rho V^2 S_w b C_l \\ M=\dfrac{1}{2}\rho V^2 S_w \bar{c} C_m \\ N=\dfrac{1}{2}\rho V^2 S_w b C_n \end{cases} \tag{3}$$

式中：ρ为空气密度；V为舰载机的空速；C_L、C_D、C_Y分别为舰载机的总升力系数、总阻力系数、总侧力系数；C_l、C_m、C_n分别为舰载机的总滚转力矩系数、总俯仰力矩系数、总偏航力矩系数。气动力系数和气动力矩系数如下：

$$\begin{cases} C_L=C_{L_0}+C_{L_\alpha}\alpha+C_{L_{\delta_e}}\delta_e+C_{L_{\delta_f}}\delta_f \\ C_D=C_{D_0}+C_{D_\alpha}\alpha+C_{D_{\delta_e}}\delta_e+C_{D_{\delta_f}}\delta_f \\ C_Y=C_{Y_\beta}\beta+C_{Y_{\delta_r}}\delta_r+C_{Y_{\delta_a}}\delta_a \end{cases} \tag{4}$$

$$\begin{cases} C_l=C_{l_\beta}\beta+C_{l_{\delta_a}}\delta_a+C_{l_{\delta_r}}\delta_r+\dfrac{b}{2V}C_{l_p}p+\dfrac{b}{2V}C_{l_r}r \\ C_m=C_{m_0}+C_{m_\alpha}\alpha+C_{m_{\delta_e}}\delta_e+C_{m_{\delta_f}}\delta_f+\dfrac{\bar{c}}{2V}C_{m_q}q \\ C_n=C_{n_\beta}\beta+C_{n_{\delta_a}}\delta_a+C_{n_{\delta_r}}\delta_r+\dfrac{b}{2V}C_{n_p}p+\dfrac{b}{2V}C_{n_r}r \end{cases} \tag{5}$$

式中：p、q、r分别为滚转角速度、俯仰角速度、偏航角速度；δ_a、δ_e、δ_r、δ_f分别为副翼、升降舵、方向舵、襟翼的舵偏角，气动导数由文献[18]给出，不再赘述。

假设飞机受到的合力转化到机体轴之后，分量为F_x、F_y、F_z，得到力方程组为

$$\begin{cases} \dot{u}=vr-wq-g\sin\theta+\dfrac{F_x}{m} \\ \dot{v}=wq-ru+g\cos\theta\sin\phi+\dfrac{F_y}{m} \\ \dot{w}=uq-pv+g\cos\theta\cos\phi+\dfrac{F_z}{m} \end{cases} \tag{6}$$

式中：u、v、w为3轴机体轴速度分量；m为舰载机的质量。

由于舰载机关于$O_bx_bz_b$平面对称，$I_{xy}=I_{yx}=I_{yz}=I_{zy}=0$，其他转动惯量为固定常量，已在表1中给出，可以得到舰载机的角动力学模型如下：

$$\begin{cases} \dot{p}=(c_1r+c_2p)q+c_3\bar{L}+c_4N \\ \dot{q}=c_5pr-c_6\left(p^2-r^2\right)q+c_7M \\ \dot{r}=(c_8p+c_2r)q+c_4\bar{L}+c_9N \end{cases} \tag{7}$$

式中：$c_1=\left(I_{yy}-I_{zz}\right)I_{zz}-I_{xz}^2\Big/\sum$；$c_2=\left(I_{xx}-I_{yy}+I_{zz}\right)I_{xz}\Big/\sum$；$c_3=I_{zz}\Big/\sum$；$c_4=I_{xz}\Big/\sum$；$c_5=I_{zz}-I_{xx}/I_{yy}$；$c_6=I_{xz}/I_{yy}$；$c_7=1/I_{yy}$；$c_8=I_{xx}\left(I_{xx}-I_{yy}\right)+I_{xz}^2\Big/\sum$；$c_9=I_{xx}\Big/\sum$，$\sum=I_{xx}I_{zz}-I_{xz}^2$。

根据舰载机的3个角速度分量(p,q,r)及其姿态角速度的关系，可以得到其运动方程组：

$$\begin{cases} \dot{\phi}=p+(r\cos\phi+q\sin\phi)\tan\theta \\ \dot{\theta}=q\cos\phi-r\sin\phi \\ \dot{\psi}=\dfrac{1}{\cos\theta}(r\cos\phi+q\sin\phi) \end{cases} \tag{8}$$

式中：$\dot{\psi}$、$\dot{\theta}$、$\dot{\phi}$为舰载机的姿态角速度。

将舰载机的机体坐标系中的状态量转化到气流坐标系中，可以得到舰载机相对气流的速度和角

度状态，如下：

$$\begin{cases} V=\sqrt{u^2+v^2+w^2} \\ \beta=\arcsin(v/V) \\ \alpha=\arctan(w/u) \end{cases} \tag{9}$$

将舰载机在机体坐标系中的线运动转化到地面坐标系中，可以得到其导航方程组，如下：

$$\begin{cases} \dot{x}_g=u\cos\theta\sin\psi+v(\sin\phi\sin\theta\sin\psi+\cos\phi\cos\psi)+ \\ \qquad w(-\sin\phi\cos\psi+\cos\phi\sin\theta\sin\psi) \\ \dot{y}_g=u\cos\theta\sin\psi+v(\sin\phi\sin\theta\sin\psi+\cos\phi\cos\psi)+ \\ \qquad w(-\sin\phi\cos\psi+\cos\phi\sin\theta\sin\psi) \\ \dot{h}=u\sin\theta-v\sin\phi\cos\theta-w\cos\phi\cos\theta \end{cases} \tag{10}$$

式中：$\dot{x}_g$、$\dot{y}_g$、$\dot{h}$为地面坐标系 3 个轴上的线速度。

根据上述分析，可以确定舰载机的纵向控制输入为 $u=[\delta_e,\delta_f,\delta_T]^T$，纵向状态向量为 $x=[V,\alpha,q,\theta,h,\gamma,\dot{\gamma}]^T$。

1.2　甲板运动模型建立

为使后续的着舰仿真更加真实，需要建立合理的甲板运动模型。本文采用文献 [19] 提出的基于 Conolly 线性理论的甲板运动模型。由于航母尺寸和质量很大，且舰载机着舰时间较短，甲板运动可以看作一个平稳的随机过程。考虑俯仰、滚转、沉浮运动对着舰的影响，将白噪声通过成型滤波器生成 3 种运动模型，如下：

$$\begin{cases} \dfrac{h_S}{N}=\dfrac{1.16s(s+0.04)}{(s^2+0.16s+0.16)(s^2+0.22s+0.3025)} \\ \dfrac{\theta_S}{N}=\dfrac{0.334\,059s}{(s^2+0.22s+0.302\,5)(s^2+0.384s+0.409\,6)} \\ \dfrac{\varphi_S}{N}=\dfrac{0.238\,350s}{(s^2+0.088\,8s+0.139\,6)(s^2+0.12s+0.25)} \end{cases} \tag{11}$$

式中：h_S、θ_S、ϕ_S分别为甲板的沉浮、俯仰、滚转运动；N为白噪声。

根据式 (11)，设定船速为 15.43 m/s 可以得到甲板 3 种运动的仿真波形，分别如图 1～图 3 所示。

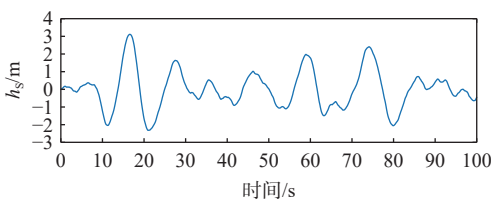

图 1　甲板沉浮运动波形

Fig. 1　Deck ups and downs movement

垂直方向上的高度变化是影响着舰点误差的主要因素，图 4 为理想着舰点垂直高度的变化（将俯仰、滚转运动全部折算成理想着舰点的高度变化）。

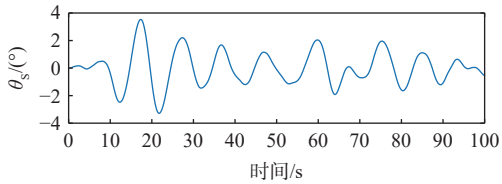

图 2　甲板俯仰运动波形

Fig. 2　Deck pitch movement

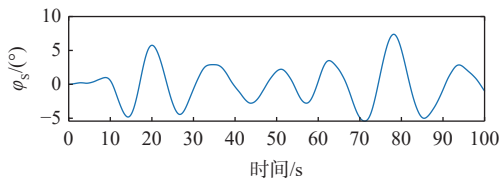

图 3　甲板滚转运动波形

Fig. 3　Deck roll movement

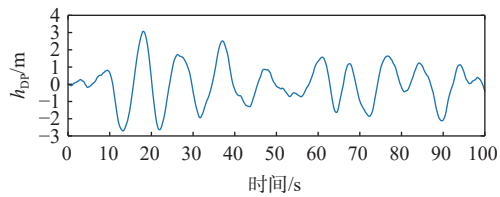

图 4　理想着舰点垂直高度变化波形

Fig. 4　Vertical height variation of ideal landing Point

为减小由甲板运动引起的着舰误差，在距舰 800 m 时需要引入甲板运动补偿器。甲板运动补偿器是一个超前滞后网络，能够补偿系统在 $\omega_s=0.2\sim0.7$ rad/s频段的相位滞后和幅值衰减，使飞机能够实时跟踪理想着舰点的垂直高度变化。甲板运动补偿器为

$$G_{DMC}(s)=66\times\left[\frac{s+3}{s+7}\right]\left[\frac{0.2s^2+0.7s+1}{s^3+9s^2+27s+27}\right] \tag{12}$$

1.3　舰尾流模型建立

本文采用文献 [20] 中的舰尾流模型，根据 MIL-F-8785C 军用规范，舰载机在x方向上距舰大约 800 m 时，开始受到舰尾流干扰，舰尾流主要由 4 个分量组成：①自由大气紊流分量u_1、v_1、w_1；②雄鸡尾流u_2、w_2；③尾流的周期性分量u_3、w_3；④尾流的随机分量u_4、v_4、w_4。

假设u_g、v_g、w_g分别表示 3 轴方向上受到的舰尾气流，表示为

$$\begin{cases} u_g=u_1+u_2+u_3+u_4 \\ v_g=v_1+v_4 \\ w_g=w_1+w_2+w_3+w_4 \end{cases} \tag{13}$$

根据舰尾流模型，设定航母船速 15.43 m/s，得到 3 级海况下，舰尾流的仿真波形，如图 5 所示。

舰尾流对舰载机着舰过程的影响主要体现在空速、迎角以及侧滑角上，从而对舰载机的姿态和

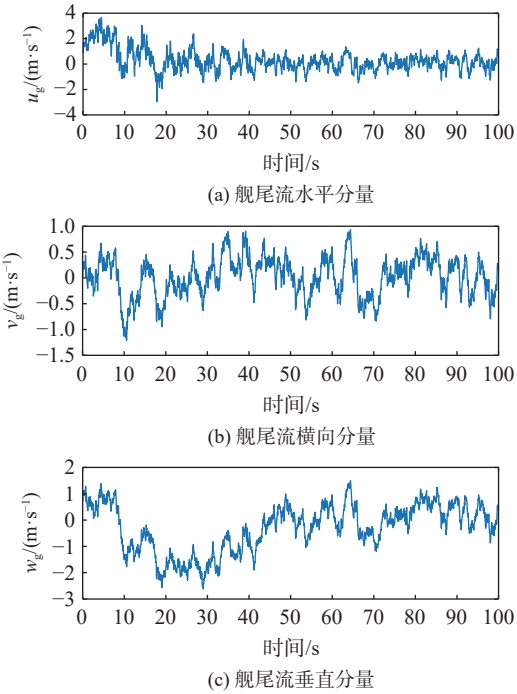

图5 舰尾流波形

Fig. 5 Airwake of the ship

航迹产生影响。将舰尾流3轴分量$[u_g, v_g, w_g]^T$经过多次坐标转换，将其转换到机体坐标系中，得到$[u'_g, v'_g, w'_g]^T$，将其与无扰动情况下的飞机线速度$[u, v, w]^T$相加得到舰尾流扰动下的机体速度$[u_{bg}, v_{bg}, w_{bg}]^T$，那么舰尾流扰动下的空速、迎角以及侧滑角如下：

$$\begin{cases} V' = \sqrt{u_{bg}^2 + v_{bg}^2 + w_{bg}^2} \\ \beta' = \arcsin\left(v_{bg} / V'\right) \\ \alpha' = \arctan\left(w_{bg} / u_{bg}\right) \end{cases} \tag{14}$$

2 PPO算法

PPO算法基于Actor-Critic框架，Actor网络输出动作而Critic网络输出状态价值函数。该算法在一块信任区域内更新策略以保证策略性能的安全性。PPO算法作为一种连续状态、连续动作空间的算法，对于舰载机自动着舰的直接升力控制是很适用的。

假设θ_{rl}表示策略$\pi_{\theta_{rl}}$的参数，PPO算法的优化目标为

$$L_{\theta_{rl}}(\theta'_{rl}) = -E_{\pi_{\theta'_{rl}}}\left[\sum_{t=0}^{\infty}\gamma_{rl}^t(\gamma_{rl}V^{\pi_{\theta_{rl}}}(s_{t+1}) - V^{\pi_{\theta_{rl}}}(s_t))\right] + E_{s_t\sim\nu^{\pi_{\theta'_{rl}}}}E_{a\sim\pi_{\theta'_{rl}}(\cdot|s_t)}\left[\frac{\pi_{\theta'_{rl}}(a|s)}{\pi_{\theta_{rl}}(a|s)}A^{\pi_{\theta_{rl}}}(s,a)\right] \tag{15}$$

式中：θ'_{rl}为相对于θ_{rl}来说更好的参数；s为某一状态；a为某一状态下采取的动作；γ_{rl}为折扣因子；$\nu^{\pi}(s) = (1-\gamma_{rl})\sum_{t=0}^{\infty}\gamma_{rl}^t P_t^{\pi}(s)$为状态访问分布，$1-\gamma_{rl}$为用来使概率加和为1的归一化因子，$P_t^{\pi}(s)$为采取策略$\pi$使得智能体在$t$时刻状态为$s$的概率。

PPO算法对上述优化目标采用PPO-截断的方式，在目标函数中进行限制，以保证新旧参数的差距不大，即

$$\arg\max E_{s_t\sim\nu^{\pi_{\theta_k}}}E_{a\sim\pi_{\theta_k}(\cdot|s_t)}\left[\min\left(\frac{\pi_{\theta'_{rl}}(a|s)}{\pi_{\theta_{rl}}(a|s)}A^{\pi_{\theta_k}}(s,a),\ \mathrm{clip}\left(\frac{\pi_{\theta'_{rl}}(a|s)}{\pi_{\theta_{rl}}(a|s)}, 1-\varepsilon, 1+\varepsilon\right)A^{\pi_{\theta_k}}(s,a)\right)\right] \tag{16}$$

式中：$\mathrm{clip}(x,l,u)$表示把x限制在$[l,u]$区间内。式(16)中的ε为一个超参数，表示PPO-截断的范围。

优势函数A_t通过广义优势估计(generalized advantage estimation, GAE)的方法进行计算，优势函数的表达式为

$$A_t = \delta_t + (\lambda\gamma_{rl})\delta_{t+1} + \cdots + (\lambda\gamma_{rl})^k\delta_{t+k} \tag{17}$$

式中：$\delta_t = r_t + \gamma_{rl}V(s_{t+1}) - V(s_t)$为时序差分误差，$V(s_t)$为Critic网络输出的状态价值函数。

PPO算法的伪代码如下：

for episode = 1 → M do:

for actor = 1 → N do:

使用策略$\pi_{\theta_{rl}}$在环境中执行k个时间步

计算优势函数$A_1, A_2, \cdots, A_k$

end for

计算目标函数$L_{\theta_{rl}}(\theta'_{rl})$

更新网络参数$\theta_{rl} \leftarrow \theta'_{rl}$

end for

3 纵向自动着舰控制器设计

舰载机自动着舰采用飞行轨迹增量控制模式。该模式通过控制飞机的航迹倾斜角来改变飞行轨迹，动态响应更快，有利于航迹的精确控制。

3.1 自动着舰期望高度指令的求解

首先设定飞机初始位置到理想着舰点的距离$X_0 = -4\,578.30$ m，对舰下滑道是一条从理想着舰点发出的射线，其角度规定为$\gamma = -3.5°$，初始高度为

$$H_0 = -X_0\tan(-\gamma/57.3) \tag{18}$$

求得初始高度$H_0 = 280$ m。假设无甲板运动的情况下，飞机到理想着舰点x方向的剩余距离为$R_X(t)$，舰载机已经飞过的横向距离为$X_f(t)$，航母行进的距离为$X_S(t)$，那么$R_X(t)$的表达式为

$$R_X(t) = -X_0 - X_f(t) + X_S(t) \tag{19}$$

根据横向剩余距离可以求解舰载机飞行基准

轨迹，如下：

$$h_{\mathrm{REF}}(t)=R_{\mathrm{X}}(t)\cdot\tan(-\gamma) \tag{20}$$

理想着舰点受甲板运动影响，在垂直方向上运动的波形$h_{\mathrm{DP}}(t)$前文已经给出，假设$h_{\mathrm{DP}}(t)$经过甲板运动补偿器之后为$h_{\mathrm{DMC}}(t)$，那么期望高度指令的表达式为

$$H_{\mathrm{c}}(t)=\begin{cases}h_{\mathrm{REF}}(t) & R_{\mathrm{X}}(t)>800\\ h_{\mathrm{REF}}(t)+h_{\mathrm{DMC}}(t) & R_{\mathrm{X}}(t)\leqslant 800\end{cases} \tag{21}$$

3.2　自动着舰基准航迹倾斜角指令的求解

基准航迹倾斜角指令通过几何计算得到。其不同于对舰下滑道的倾斜角。在航母前进过程中，对舰下滑道会随之移动，导致舰载机的实际航迹倾斜角小于对舰下滑道的夹角，几何关系如图 6 所示。

V_{S}表示航母行驶速度矢量，ψ_{d}表示斜甲板的角度，$V_{\mathrm{S}}\cos\psi_{\mathrm{d}}$表示航母行驶速度在斜甲板上的分量。$V_{\mathrm{p}}$为飞机地速，$\gamma_{\mathrm{ref}}$为飞机的基准航迹倾斜角，$\gamma$为对舰下滑道的夹角。根据图 6 的三角关系，可得

$$\frac{\sin(\gamma+\gamma_{\mathrm{ref}})}{\sin\gamma}=\frac{V_{\mathrm{S}}\cos\psi_{\mathrm{d}}}{V_{\mathrm{p}}} \tag{22}$$

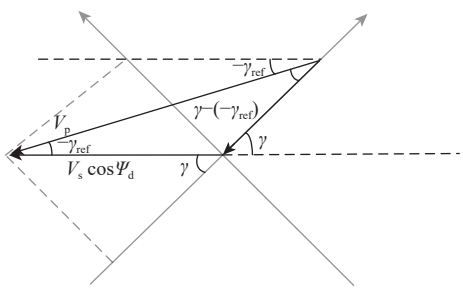

图 6　航迹倾斜角参考指令求解

Fig. 6　The solution graph for γ_{ref}

舰载机在自动着舰状态下，对舰下滑道的夹角γ取为−3.5°，由于γ和γ_{ref}都很小，有

$$\sin(\gamma+\gamma_{\mathrm{ref}})\approx\gamma+\gamma_{\mathrm{ref}}\quad\sin\gamma\approx\gamma \tag{23}$$

代入式 (22)，可得

$$\gamma_{\mathrm{ref}}=\frac{V_{\mathrm{S}}\cos\psi_{\mathrm{d}}-V_{\mathrm{p}}}{V_{\mathrm{p}}}\gamma \tag{24}$$

因此，在飞机地速$V_{\mathrm{p}}=70.016\ 8$ m/s，对舰下滑道的夹角$\gamma=-3.5°$，船速$V_{\mathrm{S}}=15$ m/s，$\psi_{\mathrm{d}}=11°$时，通过式 (24) 可以计算出基准航迹倾斜角指令$\gamma_{\mathrm{ref}}=-2.75°$。

3.3　基于 PID 的直接升力控制器设计

基于 PID 的飞行轨迹增量控制模式的框图如图 7 所示。舰载机通过襟翼的偏转产生直接升力，可以迅速改变航迹倾斜角，实现对高度指令的跟踪。

这里需要注意的是，当襟翼偏转时，会产生俯仰力矩，因此需要升降舵来平衡俯仰力矩，同时保持迎角恒定。

图 7 中，控制系统首先根据高度误差计算航迹倾斜角增量指令，并与航迹倾斜角参考指令进行比较，然后通过控制器得到期望航迹倾斜角速率，再与实际航迹倾斜角速率作差后经过控制器得到襟翼舵偏值。

3.4　基于深度强化学习的直接升力控制器设计

为了降低控制器参数调节与优化的难度以及对数学模型精度的依赖性，本文提出了一种基于 PPO 算法的纵向直接升力控制器，控制框架如图 8 所示。

将舰载机看作待训练的智能体，其交互环境包括甲板运动、舰尾流等，飞机模型以及交互环境均在 MATLAB/Simulink 中搭建。

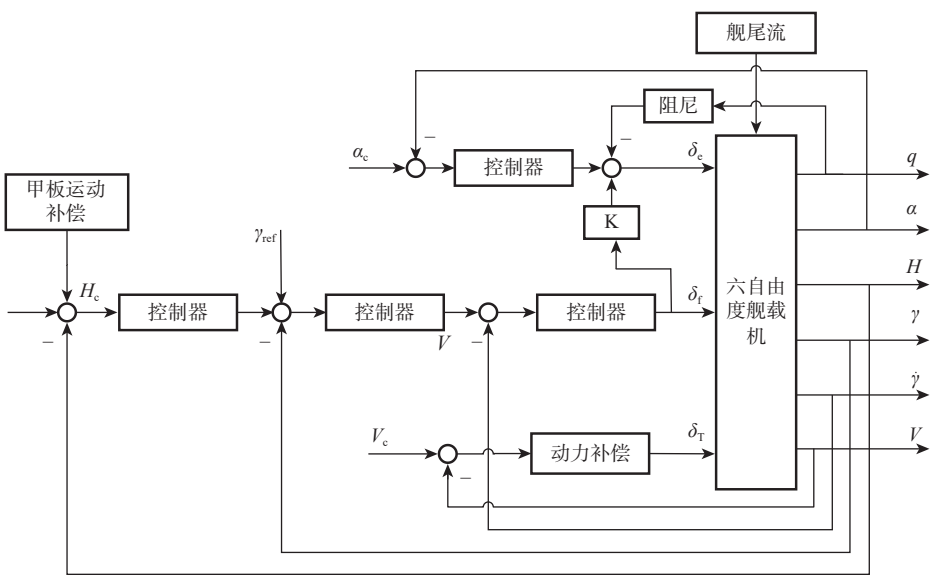

图 7　基于 PID 的直接升力控制框图

Fig. 7　Traditional direct lift control block diagram

在强化学习理论中，舰载机在着舰过程中的状态转移可以看作马尔可夫决策过程(Markov decision process, MDP)，即舰载机从一种状态转移到下一种状态的概率只与当前状态和所采取的动作有关，与之前的历史状态和动作无关，如下：

$$P(S_{t+1}) = E[S_{t+1} = s_{t+1} | S_t = s_t, A_t = a_t] \tag{25}$$

式中：s_t为飞机某一时刻的状态；s_{t+1}为飞机下一时刻的状态；a_t为飞机某一时刻采取的动作。飞机在t时刻采取动作a_t后会获得相应的奖励，如下：

$$r_t = r(s_t, a_t) \tag{26}$$

根据PPO算法的流程，飞机智能体与着舰环境交互多次，根据当前状态s_t采取动作a_t，获得奖励r_t并且状态转移至s_{t+1}，将元组(s_t, a_t, r_t, s_{t+1})存放在Mini-Batch中，当元组个数达到一定数量时，批量计算A_t以及目标函数$L_{\theta_{\mathrm{n}}}(\theta'_{\mathrm{n}})$，并更新网络参数。基于强化学习的直接升力控制器训练框图如图9所示。

3.4.1 状态-动作空间设计

本文选取一个6维的状态空间，如下：

$$\boldsymbol{S} = (H, H_{\mathrm{e}}, \theta, \gamma, \dot{\gamma}, q) \tag{27}$$

式中：H为舰载机的高度；$H_{\mathrm{e}} = H_{\mathrm{c}} - H$为实际高度与期望下滑道的纵向偏差；$\theta$为舰载机的俯仰角；$\gamma$为舰载机的航迹倾斜角；$\dot{\gamma}$为舰载机的航机倾斜角速率；$q$为舰载机的俯仰角速度。

动作空间的维度为1，为飞机襟翼的偏转，如下：

$$\boldsymbol{a} = (\delta_{\mathrm{f}}) \tag{28}$$

3.4.2 奖励函数设计

奖励函数在智能体训练的过程中起着至关重要的作用，合理的奖励函数可以有效地引导智能体进行学习。

奖励函数分为连续型和离散型，连续型奖励函数计算量较大但可以提高训练过程的收敛性，对网

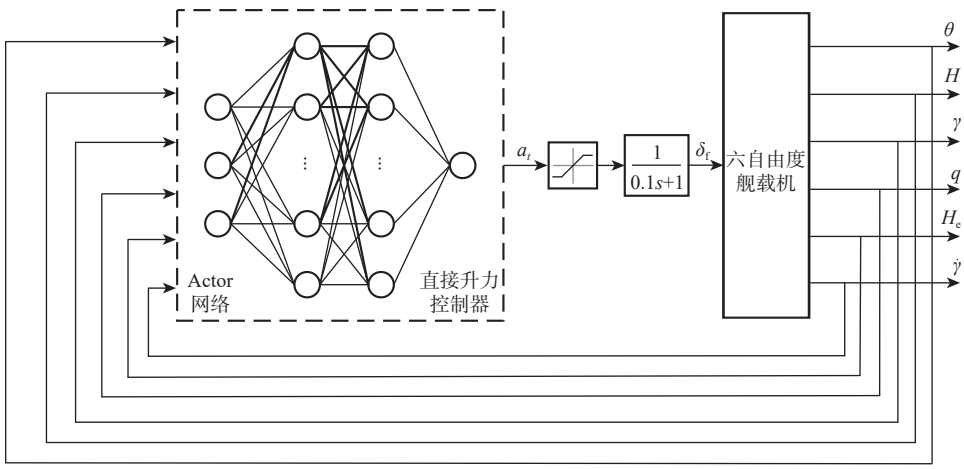

图8 基于深度强化学习的直接升力控制框图

Fig. 8 Deep Reinforcement Learning control block diagram

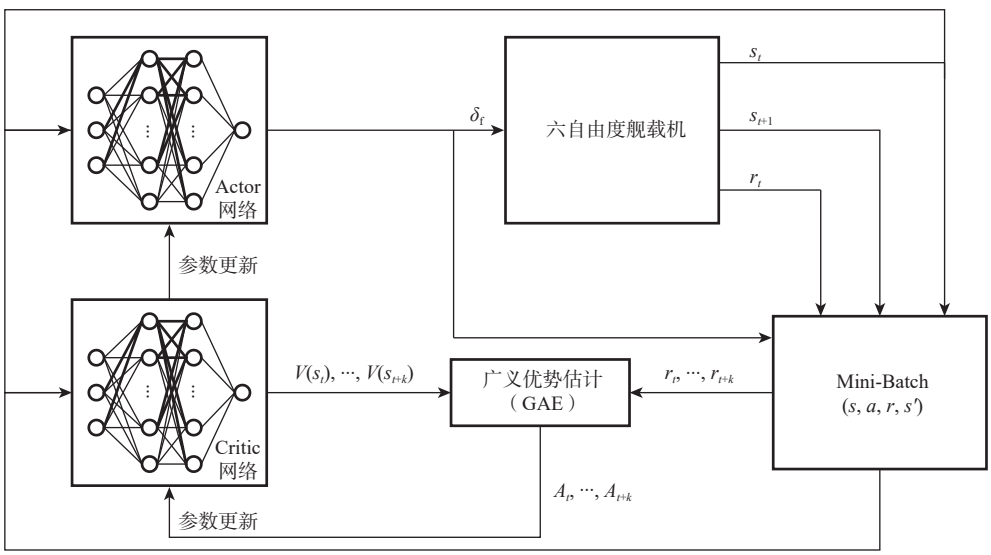

图9 控制器训练框图

Fig. 9 Controller training block diagram

络结构的复杂度较低。而离散型奖励函数由于变化的离散性，使得收敛速度变慢但可以引导智能体在状态空间中找到更好的奖励区域。基于此，本文采用混合型奖励函数，离散函数用于驱动智能体远离不良状态，连续函数通过在目标状态附近提供平滑奖励来提高收敛性。

还有一点值得注意的是，在训练过程中，智能体只会学习如何最大化收益，如果想让其完成某些指定任务，就必须将目标融入到奖励函数中。

高度误差的奖励函数为

$$R_{H_e}=\begin{cases}-0.2H_e^2 & |H_e|>10\\ 63.25\sqrt{\dfrac{1}{|H_e|}} & 0.04<|H_e|\leqslant 10\\ 400 & |H_e|\leqslant 0.04\end{cases} \tag{29}$$

舵面偏转的奖励函数为

$$R_{\delta_f}=-0.5\delta_f^2 \tag{30}$$

航迹倾斜角的奖励函数为

$$R_\gamma=\begin{cases}-(\gamma_{ref}-\gamma)^2 & |\gamma_e|<0.2\\ 15 & |\gamma_e|\geqslant 0.2\end{cases} \tag{31}$$

式中：$\gamma_e=\gamma_{ref}-\gamma$。

俯仰角速度的奖励函数为

$$R_q=-0.2(0-q)^2 \tag{32}$$

训练步进奖励为

$$R_{step}=30 \tag{33}$$

终止训练奖励函数为

$$R_{final}=\begin{cases}10\,000 & H\leqslant 0,0\leqslant H_e<1.8\\ 10\,000 & H_c\leqslant 0,-1\leqslant H_e\leqslant 0\\ -30\,000 & H\leqslant 0,H_e\geqslant 1.8\\ -30\,000 & H_c\leqslant 0,H_e<-1\\ -30\,000 & \theta>20\text{或}\theta<-20\\ -30\,000 & |H_e|\geqslant 25\\ -30\,000 & H>H_0+5\\ -30\,000 & H\leqslant 0,(|H_d|\geqslant 5\text{或}(\theta>10\text{或}\theta\leqslant 0))\end{cases} \tag{34}$$

式中：H_c为期望高度；H_d为舰载机的下沉率；H_0为着舰的初始高度。

总奖励函数的表达式为

$$R_{all}=R_{H_e}+R_{\delta_f}+R_\gamma+R_q+R_{step}+R_{final} \tag{35}$$

3.4.3　超参数设计

PPO 算法中 Actor-Critic 网络的结构以及超参数设计如表 2 所示。

Actor 网络与 Critic 网络均采用全连接结构，隐含层激活用函数为 ReLu 函数。Actor 网络中，动作均值的激活函数为 tanh 函数，动作方差的激活函数为 Softplus 函数，根据概率分布进行高斯采样得到具体的舵面偏转值。Loss function 均采用 Adam 优化器进行梯度更新。

表 2　算法参数设置

Table 2　PPO algorithm parameter setting table

参数	数值
PyCharm平台IP端口	127.0.0.1.500 00(地址)
Simulink平台IP端口	127.0.0.1.500 01(地址)
Actor网络结构	6×128×128×2
Critic网络结构	6×128×128×128×1
Actor网络学习率	1×10^{-3}
Critic网络学习率	5×10^{-3}
回合数	20 000
步数	1 500
γ_{rl}	0.99
λ	0.95
ε	0.2
训练前交互步数	2 048
交叉熵损失函数权重	0.01
小批次数据量	256
平均奖励计算回合数	10
动作界限值	30

3.4.4　控制律设计

假设 Actor 网络输出的动作均值为μ_f，方差为σ_f^2。均值与方差均通过 Actor 网络计算得出。Actor 网络的输入为 6 个状态的列向量$\boldsymbol{x}=\left[H,H_e,\theta,\gamma,\dot{\gamma},q\right]^{\mathrm{T}}$，经过 2 个隐含层之后得到输出$\boldsymbol{y}=[\mu_f,\sigma_f^2]^{\mathrm{T}}$，约定：

1）变量z_i^l为第l层的第i个神经元，变量a_i^l为经过激活函数后的第l层的第i个神经元。

2）$\boldsymbol{W}^l$表示权重矩阵，权重$w_{i,j}^l$表示第$l-1$层的第j个神经元与第l层第i个神经元之间的权重。

Actor 网络第 1 个隐含层的输入$\boldsymbol{z}^1=[z_1^1,z_2^1,\cdots,z_{127}^1,z_{128}^1]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{z}^1=\boldsymbol{W}^1\boldsymbol{x}=\begin{cases}z_1^1=w_{1,1}^1H+w_{1,2}^1H_e+w_{1,3}^1\theta+w_{1,4}^1\gamma+\\ \qquad w_{1,5}^1\dot{\gamma}+w_{1,6}^1q\\ z_2^1=w_{2,1}^1H+w_{2,2}^1H_e+w_{2,3}^1\theta+w_{2,4}^1\gamma+\\ \qquad w_{2,5}^1\dot{\gamma}+w_{2,6}^1q\\ \vdots\\ z_{128}^1=w_{128,1}^1H+w_{128,2}^1H_e+w_{128,3}^1\theta+\\ \qquad w_{128,4}^1\gamma+w_{128,5}^1\dot{\gamma}+w_{128,6}^1q\end{cases} \tag{36}$$

Actor 网络第 1 个隐含层的输出$\boldsymbol{a}^1=[a_1^1,a_2^1,\cdots,a_{127}^1,a_{128}^1]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{a}^1=f_{\mathrm{ReLU}}(\boldsymbol{z}^1)=\begin{cases}a_1^1=f_{\mathrm{ReLU}}(z_1^1)\\ a_2^1=f_{\mathrm{ReLU}}(z_2^1)\\ \vdots\\ a_{128}^1=f_{\mathrm{ReLU}}(z_{128}^1)\end{cases} \tag{37}$$

Actor网络第2个隐含层的输入$\boldsymbol{z}^2=[z_1^2,z_2^2,\cdots,z_{127}^2,z_{128}^2]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{z}^2=\boldsymbol{W}^2\boldsymbol{a}^1=\begin{cases}w_{1,1}^2a_1^1+w_{1,2}^2a_2^1+\cdots+w_{1,127}^2a_{127}^1+\\ \quad w_{1,128}^2a_{128}^1\\ w_{2,1}^2a_1^1+w_{2,2}^2a_2^1+\cdots+w_{2,127}^2a_{127}^1+\\ \quad w_{2,128}^2a_{128}^1\\ \vdots\\ w_{128,1}^2a_1^1+w_{128,2}^2a_2^1+\cdots+w_{128,127}^2a_{127}^1+\\ \quad w_{128,128}^2a_{128}^1\end{cases} \tag{38}$$

Actor网络第2个隐含层的输出$\boldsymbol{a}^2=[a_1^2,a_2^2,\cdots,a_{127}^2,a_{128}^2]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{a}^2=f_{\mathrm{ReLU}}(\boldsymbol{z}^2)=\begin{cases}a_1^2=f_{\mathrm{ReLU}}(z_1^2)\\ a_2^2=f_{\mathrm{ReLU}}(z_2^2)\\ \vdots\\ a_{128}^2=f_{\mathrm{ReLU}}(z_{128}^2)\end{cases} \tag{39}$$

Actor网络输出层的输入$\boldsymbol{z}^3=[z_1^3,z_2^3]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{z}^3=\boldsymbol{W}^3\boldsymbol{a}^2=\begin{cases}w_{1,1}^3a_1^2+w_{1,2}^3a_2^2+\cdots+w_{1,127}^3a_{127}^2+w_{1,128}^3a_{128}^2\\ w_{2,1}^3a_1^2+w_{2,2}^3a_2^2+\cdots+w_{2,127}^3a_{127}^2+w_{2,128}^3a_{128}^2\end{cases} \tag{40}$$

Actor网络输出层的输出$\boldsymbol{y}=[\mu_{\mathrm{f}},\sigma_{\mathrm{f}}^2]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{y}=\begin{bmatrix}\mu_{\mathrm{f}}\\ \sigma_{\mathrm{f}}^2\end{bmatrix}=\begin{cases}f_{\tanh}(z_1^3)\\ f_{\mathrm{Softplus}}(z_2^3)\end{cases} \tag{41}$$

在某一时刻，智能体的状态到达s，此状态下动作a服从均值为μ_{f}、方差为σ_{f}^2的正态分布：

$$a\sim N(\mu_{\mathrm{f}},\sigma_{\mathrm{f}}^2) \tag{42}$$

则动作a的概率密度为

$$f(a)=\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{f}}}\mathrm{e}^{-\frac{(a-\mu_{\mathrm{f}})^2}{2\sigma_{\mathrm{f}}^2}} \tag{43}$$

襟翼偏转的表达式为

$$\delta_{\mathrm{f}}=\delta_{\mathrm{f}_0}+\mathrm{sample}\left[N(\mu_{\mathrm{f}},\sigma_{\mathrm{f}}^2)\right] \tag{44}$$

式中：δ_{f_0}为襟翼配平值；sample为高斯采样函数，即依照概率进行动作的选取。也就是说，某个具体动作的概率密度越大，被选中的概率就越大。在训练结束后，方差σ_{f}^2很小，那么随机选取的动作将十分逼近均值μ_{f}。

纵向的升降舵及油门的控制仍采用经典的PID控制器，控制律为

$$\delta_{\mathrm{e}}=\delta_{\mathrm{e}_0}+K_qq+\left(K_p+\frac{1}{s}K_i\right)(\alpha_{\mathrm{c}}-\alpha)+K\Delta\delta_{\mathrm{f}} \tag{45}$$

$$\delta_{\mathrm{T}}=\delta_{\mathrm{T}_0}+\left(K_p+\frac{1}{s}K_i\right)(V_{\mathrm{c}}-V) \tag{46}$$

式中：δ_{e_0}为升降舵配平值；$K=-M_{\delta_{\mathrm{f}}}/M_{\delta_{\mathrm{e}}}$为襟翼与升降舵的解耦信号；$M_{\delta_{\mathrm{f}}}$和$M_{\delta_{\mathrm{e}}}$分别为襟翼和升降舵的俯仰力矩系数；$\delta_{\mathrm{T}_0}$为油门开度的配平值。

4 仿真验证

本文的环境在Simulink中搭建，包括舰载机的动力学/运动学模型、着舰引导律、甲板运动模型、舰尾流模型等，飞机参数使用F/A-18的数据，训练框架使用Python3.8.10与PyTorch1.13.0。基于深度强化学习的直接升力控制器的回合奖励曲线如图10所示。

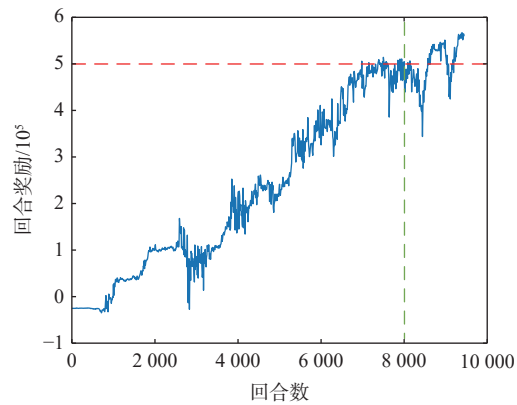

图10 回合奖励收敛曲线

Fig. 10 Episode reward curve

从图10中可以看出，训练8 000回合后，回合奖励基本已经维持在500 000以上并逐渐开始收敛，在9 440回合左右，达到540 000的收敛均值。控制器的平均奖励曲线如图11所示。

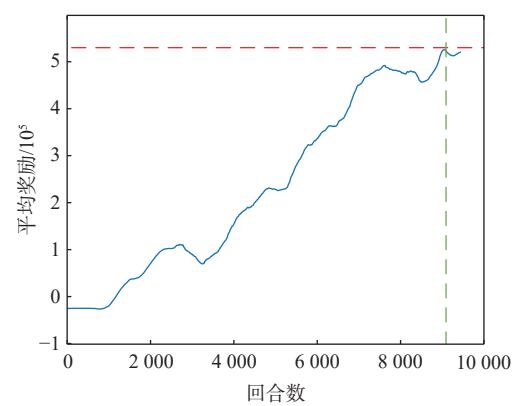

图11 平均奖励收敛曲线

Fig. 11 Average reward curve

从图11中可以看出，训练到9 440回合时，奖励均值(取训练过程中最新50个episodes的奖励求平均)已经达到540 000的收敛值。图12为下滑道跟踪曲线，图13为下滑道跟踪误差曲线。

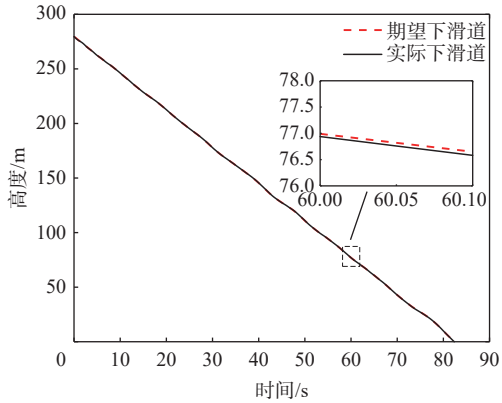

图 12　下滑道跟踪曲线

Fig. 12　Glide track curve

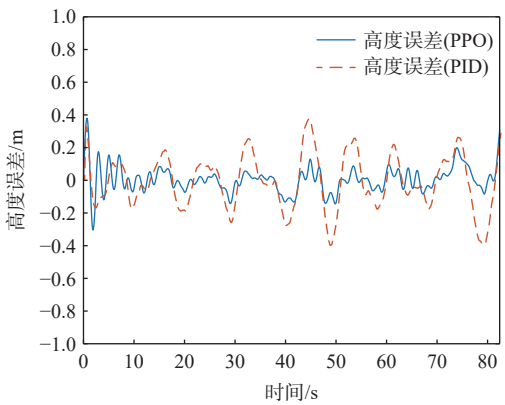

图 13　高度误差曲线

Fig. 13　Height error curve

本文设置的着舰的初始高度 $H_0 = 280$ m，图 12 中红色虚线为期望下滑道，黑色实线为实际下滑道。根据图中的细节可以看出在 PPO 控制器的作用下，下滑道有 0.05 m 左右的误差。通过图 13 的对比可以明显看出，PPO 控制器作用下的高度误差比 PID 控制器更小，对于着舰过程中舰尾流以及甲板运动的影响有更强的抑制作用。图 14 为 PPO 控制器作用下的襟翼偏转曲线图。

本文采用的是增量控制，即舵偏值为减去配平数值后的大小。从图 14 中可以看出，对于 PPO 控制器，开始下滑的前 10 s 为了快速跟踪标准下滑道，襟翼偏转角度较大，达到 20°。10 s 之后趋于稳定，在 0 值附近有一些微调。对于 PID 控制器，稳定后也处于 0 值附近，但在着舰后期，受甲板运动与舰尾流的影响，襟翼舵偏幅度比 PPO 控制器作用下的襟翼舵偏幅度大。图 15 为不同控制器作用下的航迹倾斜角曲线。

在着舰的前 10 s，PID 控制器作用下的航迹倾斜角调节幅度比 PPO 控制器更小，10 s 之后，PID 控制器对−2.75°的角度保持效果更差。在舰载机距舰 800 m 时，受到甲板运动以及舰尾流的影响，

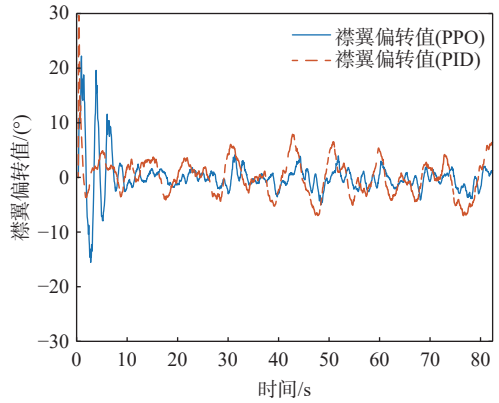

图 14　襟翼偏转曲线

Fig. 14　Flap deflection curve

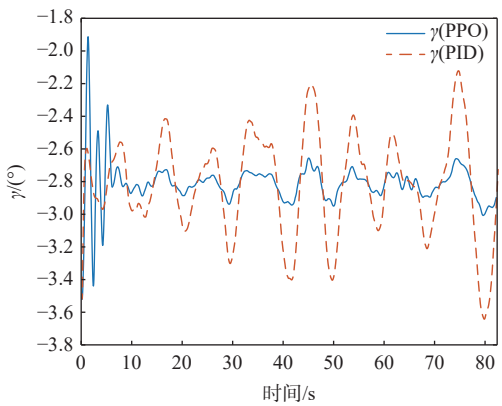

图 15　航迹倾斜角曲线

Fig. 15　Flight path angle

PID 控制器作用下的航迹倾斜角出现了更大幅度的抖动。

舰载机的迎角曲线如图 16 所示。

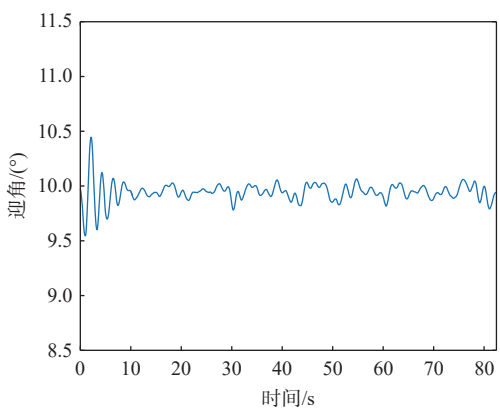

图 16　迎角曲线

Fig. 16　Attack angle curve

从图 16 中可以看出，在升降舵的作用下，迎角基本保持在 9.928 9°的期望值附近。

舰载机的空速响应如图 17 所示。

从图 17 中可以看出，在油门的控制作用下，飞机空速保持在70 m/s的期望值附近。

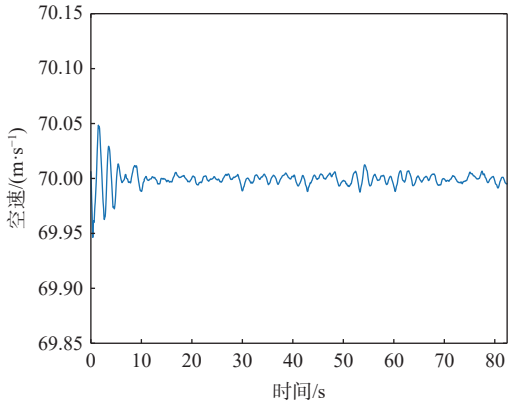

图 17 空速曲线

Fig. 17 Airspeed curve

5 结 论

1）训练收敛后的PPO控制器，实现了自动着舰的精准控制，能够较好地跟踪期望下滑道，最大跟踪误差基本控制在±0.2 m以内。

2）PPO控制器对舰尾流有更强的抑制作用，在距舰800 m，受舰尾流干扰时，航迹倾斜角的抖动幅度更小，控制在(−3,−2.6)的区间内。

3）在PPO算法的作用下，控制器的调节时间更短，从高度误差、航迹倾斜角的结果中均可看出，比起PID控制器调节时间缩短了2 s左右。

4）PPO算法作用下的襟翼舵面偏转幅度更小，稳定后舵偏角度控制在5°以内，这对舵面有更好的保护作用。

为使PPO算法对自动着舰具有更好的控制效果，仍需继续优化奖励函数以及算法的超参数，在提高训练收敛速度的同时提高控制精度。同时，可以考虑将升降舵以及油门开度的控制也替换为强化学习的方法。

参考文献（References）

[1] 张守权, 王华明. 舰载机全自动着舰综述[J]. 飞机设计, 2022, 42(3): 20-24.
ZHANG S Q, WANG H M. Summary report on automatic carrier landing system[J]. Aircraft Design, 2022, 42(3): 20-24 (in Chinese).

[2] 甄子洋, 王新华, 江驹, 等. 舰载机自动着舰引导与控制研究进展[J]. 航空学报, 2017, 38(2): 020435.
ZHEN Z Y, WANG X H, JIANG J, et al. Research progress in guidance and control of automatic carrier landing of carrier-based aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(2): 020435 (in Chinese).

[3] 张守权. 基于直接力控制的人工着舰技术综述[J]. 飞机设计, 2022, 42(2): 21-25.
ZHANG S Q. A review of manual carrier landing technology based on direct force control[J]. Aircraft Design, 2022, 42(2): 21-25 (in Chinese).

[4] WU W H, SONG L T, ZHANG Y, et al. Nonlinear comprehensive decoupling controller based on direct lift control for carrier landing[J]. IEEE Access, 2022, 10: 113875-113887.

[5] YAN Y D, YANG J, LIU C J, et al. On the actuator dynamics of dynamic control allocation for a small fixed-wing UAV with direct lift control[J]. IEEE Transactions on Control Systems Technology, 2020, 28(3): 984-991.

[6] GUAN Z Y, LIU H, ZHENG Z W, et al. Moving path following with integrated direct lift control for carrier landing[J]. Aerospace Science and Technology, 2022, 120: 107247.

[7] 罗飞, 张军红, 耿延升, 等. 动态逆反馈控制框架下直接升力控制的控制分配研究[J]. 航空科学技术, 2022, 33(8): 51-60.
LUO F, ZHANG J H, GENG Y S, et al. Study on control allocation technology of direct lift control under dynamic inversion feedback control framework[J]. Aeronautical Science & Technology, 2022, 33(8): 51-60 (in Chinese).

[8] 魏毅寅, 郝明瑞, 范宇. 人工智能技术在宽域飞行器控制中的应用[J]. 宇航学报, 2023, 44(4): 530-537.
WEI Y Y, HAO M R, FAN Y. The application of artificial intelligence technology in wide-field vehicle control[J]. Journal of Astronautics, 2023, 44(4): 530-537 (in Chinese).

[9] 孙智孝, 杨晟琦, 朴海音, 等. 未来智能空战发展综述[J]. 航空学报, 2021, 42(8): 525799.
SUN Z X, YANG S Q, PIAO H Y, et al. A survey of air combat artificial intelligence[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(8): 525799 (in Chinese).

[10] 付宇鹏, 邓向阳, 何明, 等. 基于强化学习的固定翼飞机姿态控制方法[J]. 控制与决策, 2023, 38(9): 2505-2510.
FU Y P, DENG X Y, HE M, et al. Reinforcement learning based attitude controller design[J]. Control and Decision, 2023, 38(9): 2505-2510 (in Chinese).

[11] 张瑞卿, 钟睿, 徐毅. 基于强化学习的航天器姿态控制器设计[J]. 上海航天 (中英文), 2023, 40(1): 80-85.
ZHANG R Q, ZHONG R, XU Y. Satellite attitude control based on reinforcement learning method[J]. Aerospace Shanghai (Chinese & English), 2023, 40(1): 80-85 (in Chinese).

[12] 金磊, 杨绍龙. 基于强化学习的航天器姿态预设性能容错控制[J]. 北京航空航天大学学报, 2024, 50(8): 2404-2412.
JIN L, YANG S L. Fault-tolerant control of spacecraft attitude with prescribed performance based on reinforcement learning[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(8): 2404-2412 (in Chinese).

[13] 付宇鹏, 邓向阳, 朱子强, 等. 基于模仿强化学习的固定翼飞机姿态控制器[J]. 海军航空大学学报, 2022, 37(5): 393-399.
FU Y P, DENG X Y, ZHU Z Q, et al. Imitation reinforcement learning based attitude controller for fixed-wing aircraft[J]. Journal of Naval Aviation University, 2022, 37(5): 393-399 (in Chinese).

[14] 周攀, 黄江涛, 章胜, 等. 基于深度强化学习的智能空战决策与仿真[J]. 航空学报, 2023, 44(4): 126731.
ZHOU P, HUANG J T, ZHANG S, et al. Intelligent air combat decision making and simulation based on deep reinforcement learning[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(4): 126731 (in Chinese).

[15] 黄江涛, 刘刚, 周攀, 等. 基于深度强化学习技术的舰载无人机自主着舰控制研究[J]. 南京师范大学学报 (工程技术版), 2022,

22(3): 63-71.
HUANG J T, LIU G, ZHOU P, et al. Research on autonomous landing control of carrier-borne UCAV based on deep reinforcement learning technology[J]. Journal of Nanjing Normal University (Engineering and Technology Edition), 2022, 22(3): 63-71 (in Chinese).

[16] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimizationalgorithms[J/OL].(2017-08-28)[2023-06-14].https://doi.org/10.48550/arXiv.1707.06347.

[17] GU Y, CHENG Y H, YU K, et al. Anti-martingale proximal policy optimization[J]. IEEE Transactions on Cybernetics, 2023, 53(10): 6421-6432.

[18] CHAKRABORTY A, SEILER P, BALAS G J. Susceptibility of F/A-18 flight controllers to the falling-leaf mode: linear analysis[J]. Journal of Guidance, Control, and Dynamics, 2011, 34(1): 57-72.

[19] 张永花. 舰载机着舰过程甲板运动建模及补偿技术研究[D]. 南京: 南京航空航天大学, 2012: 9-10.
ZHANG Y H. Research on deck motion modeling and compensation technology of carrier-based aircraft landing process[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012: 9-10 (in Chinese).

[20] WOODCPCK T J. Background information and user guide for MIL-F-8785C[R]. Washington, D. C. : Air Force Wright Aeronautical, 1982.

Direct lift control technology of carrier aircraft landing based on reinforcement learning

LIU Rendi，JIANG Ju*，ZHANG Zhe，LIU Xiang

(College of Automation Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China)

Abstract：The direct lift control method of automatic landing based on Proximal Policy Optimization (PPO) algorithm was proposed to solve the problem that it is easy to touch ship due to disturbance of deck movement and carrier air wake during automatic landing of carrier aircraft. The PPO controller takes six state variables of pitch angle, height, flight path angle, pitch angle rate, height error and flight path angle rate as input and output as flap deflection angle, realizing the rapid response of carrier aircraft in different landing states of flight path angle. Compared with traditional PID controller, the Actor-Critic network in PPO controller greatly improves the calculation efficiency of control quantity, and also reduces the difficulty of parameter optimization. The simulation experiment in this paper is based on the dynamics/kinematics model of F/A-18 aircraft constructed in Matlab/Simulink. The intensive learning and training environment built on PyCharm platform is used to realize the data interaction between the two platforms through user datagram protocol (UDP) communication. The simulation results show that the proposed method has the characteristics of fast response speed and small dynamic error, and can stabilize the landing height error within ±0.2 m, with high control accuracy.

Keywords：carrier aircraft landing; reinforcement learning; proximal policy optimization algorithm; direct lift control; UDP communication

Received：2023-06-21；**Accepted**：2023-11-10；**Published Online**：2023-11-24 17:51
URL：link.cnki.net/urlid/11.2625.V.20231123.1537.005

* **Corresponding author**. E-mail：jiangju@nuaa.edu.cn