文章编号:1007－6735(2026)03－0308－08　DOI: 10.13255/j.cnki.jusst.20250307004

嵌入涡驱动奖励的深度强化学习控制翼型流动分离研究

孙　瑶[1,2]，　董祥瑞[1,2]，　王　淇[1,2]，　蔡小舒[1,2]
（1. 上海理工大学 能源与动力工程学院，上海 200093；2. 上海市动力工程多相流传热重点实验室，上海 200093）

摘要：为抑制翼面分离流的产生及涡脱落的不稳定性，采用深度强化学习（DRL）策略，对弱湍流条件下翼型的流动分离问题进行优化控制。在基于软演员−评论家（SAC）算法的主动流动控制策略训练中，采用一种新型奖励函数，引入 Liutex 统计量作为流场中涡结构旋转强度与涡核大小反馈环境的关键指标，结合气动性能指标最终构成。训练结果表明，这种基于涡驱动的奖励函数在有效消除大涡和优化气动性能方面具有显著优势，与 DRL 框架相结合，可有效提升气动性能，实现阻力和升力系数波动的最小化，凸显该方法在先进流动控制策略中的应用潜力。

关键词：智能流动控制；深度强化学习；合成射流；涡驱动奖励
中图分类号：TK 83　　**文献标志码：**A

Airfoil flow separation control using deep reinforcement learning with vortex-driven reward

SUN Yao[1,2], DONG Xiangrui[1,2], WANG Qi[1,2], CAI Xiaoshu[1,2]
(1. *School of Energy and Power Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*; 2. *Shanghai Key Laboratory of Multiphase Flow and Heat Transfer for Power Engineering, Shanghai 200093, China*)

Abstract: To suppress the generation of separation flow on the airfoil surface and the instability of vortex shedding, a deep reinforcement learning (DRL) strategy was adopted to optimize the control of flow separation around an airfoil under low-turbulence conditions. Specifically, during the training of an active flow control policy based on the soft actor-critic (SAC) algorithm, a novel reward function was proposed. This reward function incorporated the Liutex statistic as a key metric to quantify the rotational intensity of vortex structures and vortex core size within the flow field, and was ultimately constructed

收稿日期: 2025−03−07
基金项目: 国家自然科学基金资助项目（52476033）
第一作者: 孙　瑶（2002—），女，硕士研究生．研究方向：空气动力学、流动控制．E-mail：242180004@st.usst.edu.cn
通信作者: 董祥瑞（1991—），女，副教授．研究方向：计算流体力学、湍流理论及流动控制．E-mail：dongxr1154@126.com

引文格式: 孙瑶，董祥瑞，王淇，等．嵌入涡驱动奖励的深度强化学习控制翼型流动分离研究[J]．上海理工大学学报，2026，48(3)：308-315.
Citation: SUN Yao, DONG Xiangrui, WANG Qi, et al. Airfoil flow separation control using deep reinforcement learning with vortex-driven reward[J]. Journal of University of Shanghai for Science and Technology, 2026, 48(3): 308-315.

by combining aerodynamic performance metrics. The training results show that the reward function based on vortex-driven has significant advantages in effectively eliminating large vortices and optimizing aerodynamic performance. The combination of the vortex-driven reward function and the DRL framework effectively improves the aerodynamic performance and minimizes fluctuations of drag and lift coefficients, demonstrating the application potential of this method in advanced flow control strategies.

Keywords: *intelligent flow control*; *deep reinforcement learning*; *synthetic jet*; *vortex-driven reward*

流动控制是流体力学的重要研究方向，在航空航天、能源动力等领域具有广泛应用。过去几十年，被动与主动流动控制技术发展迅速，尤其是主动流动控制（active flow control, AFC），因其可调节性而受到关注。流体流动的时空演化受非线性 Navier-Stokes 方程主导，呈现高维、多模态和多尺度等复杂特性[1]，传统控制技术在实时性、效率及节能等方面面临挑战。AFC 根据是否通过感知环境接收流动反馈信息，分为开环控制和闭环控制[2]。与基于预定策略的周期性信号的开环控制相比，闭环 AFC 能够根据环境的反馈调整执行器的控制策略[3]。

随着人工智能（AI）技术的发展，强化学习（reinforcement learning, RL）及深度强化学习（deep reinforcement learning, DRL）在 AFC 中得到应用。Rabault 等[4]率先在低雷诺数圆柱流动中引入 DRL，展现出较遗传算法更优的鲁棒性和学习效率[5]。DRL 通过深度神经网络（deep neural network, DNN）学习环境状态与控制动作之间的映射，近年来，深度确定性策略梯度（deep deterministic policy gradient, DDPG）[6]、近端策略优化（proximal policy optimization, PPO）[7]及软演员−评论家（soft actor-critic, SAC）算法[8]等无模型 DRL 算法相继被用于训练智能体与环境直接交互的更新策略。然而，不同算法在实际应用中各有局限。PPO 在处理复杂任务时，常因需要大量的梯度更新步数和样本量，导致采样效率较低；DDPG 虽然能在连续动作中学习确定性策略，但在高维环境下容易出现探索不足、策略收敛到局部最优以及训练不稳定等问题。相比之下，SAC 是一种基于最大熵强化学习理论的演员−评论家 DRL 算法，其核心思想是在优化策略时，最大化奖励的同时保持策略的熵值，从而增强探索能力，并提高训练的稳定性[8]。

AFC 中的执行器包括合成喷流[9]、壁面振动[10]、几何变形[11]。系统的效能与环境感知的有效性共同决定了控制效果。因此，很多研究者关注传感器布局的相关优化。例如：Paris 等[12]引入随机门控层优化探头布置；Hu 等[5]结合物理信息神经网络（physics-informed neural networks, PINN）与 DRL，实现探头数量从 164 个减少至 4 个，提高控制效果。

奖励函数设计是 DRL 控制策略的核心。Rabault 等[4]基于时间平均阻力系数定义奖励，实现 8% 阻力降低。Wang 等[13]在 PPO 框架中加入喷流能耗惩罚项，优化翼型流动控制。Viquerat 等[11]将 DRL 应用于翼型形状优化，通过设定合适的奖励函数，并结合贝塞尔曲线，无需任何先验条件即可获得类似机翼的最佳形状。Gong 等[14]在跨音速颤振控制中引入动能奖励，Guastoni 等[15]基于 DDPG 优化通道流动阻力控制，成功降低 43% 的阻力。Li 等[16]利用动力学模态分解（dynamic mode decomposition, DMD）获得涡街能量信息，提出增强稳定性的奖励函数。

综上，尽管 DRL 在流动控制中潜力巨大，但现有研究多依赖宏观气动参数或单一能量指标构建奖励函数，难以实时表征非定常涡街演化对气动性能的深层影响。针对翼型绕流控制，本研究提出一种基于涡驱动的奖励函数设计方法，通过引入流场动态特征提升策略，增强智能体对瞬态物理演化的感知精度。该方法弥补了传统奖励函数在非定常特征提取上的不足，显著提升了 DRL 在复杂绕流环境下的学习效率与控制性能。

1　问题描述及数值方法

1.1　问题描述

如图 1 所示，本研究采用一个二维计算域，模拟 NACA0012 翼型在攻角为 10°时的气流情况，图中标注了相应的边界条件。翼型的前缘位置设在原点 $O(0,0,0)$，C 表示翼型的弦长。不同 DRL 配置的算例描述列在表 1 中。算例 1 表示无控制情

况下的翼型流动，而算例 2 则采用了 DRL 控制的合成喷射控制。合成射流通过在翼型表面施加动量注入，改变流体的流动方向和速度分布。与算例 2 相比，基于涡的奖励函数在算例 3 中得到了引入并验证。由图 1 可见，翼型表面设置了 3 个合成射流孔，合成射流的净质量流量保持为 $\sum_{i=1}^{3} Q_i = 0$。其中，Q_i 表示第 i 个射流孔的质量流量，$Q_i = U_{\mathrm{jet},i} L_{\mathrm{jet}}$，$Q_3 = Q_1 + Q_2$，$L_{\mathrm{jet}}$为喷射孔的宽度，$U_{\mathrm{jet},i}$为每个喷射器的喷射速度。$U_{\mathrm{jet},i}$由 DRL 代理在其独立训练过程中驱动。为实现流畅的 AFC 过程，喷射激励经过平滑处理，从而防止升力波动过大。

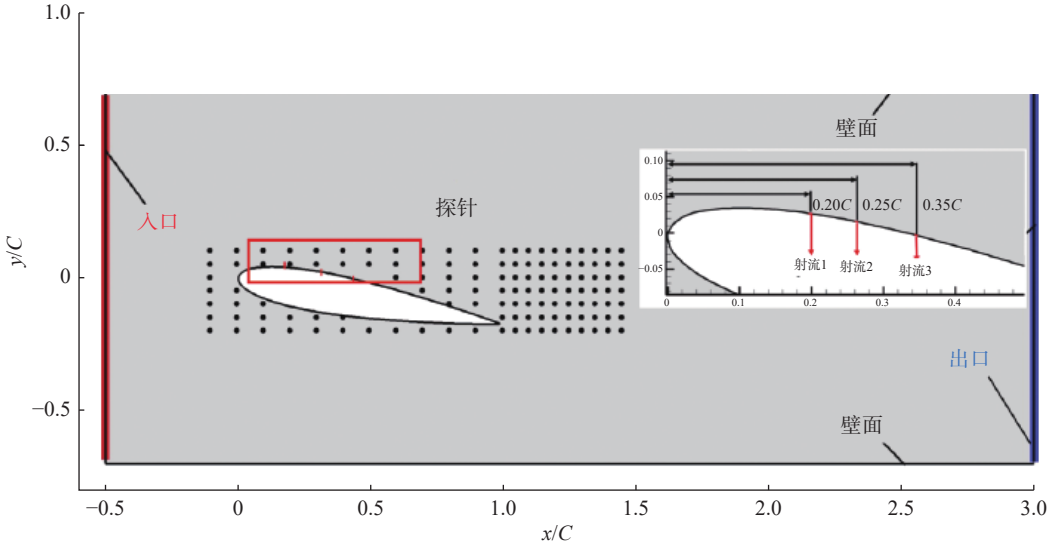


图 1 数值模拟计算区域与射流区域

Fig.1 Computation domain for numerical simulation and jet zone

表 1 不同 DRL 配置和奖励函数的算例描述

Tab.1 Case description with different DRL configuration and reward functions

算例	射流个数	探针数量	奖励函数
1	—	—	—
2	3	124	R_{c}
3	3	124	R_{v}

1.2 控制方程和数值方法

在本研究中，流动是黏性且不可压缩的，采用黏性不可压 NS 方程，其无量纲守恒形式的控制方程如下：

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot(\nabla \boldsymbol{u}) = -\nabla p + \frac{1}{Re}\Delta \boldsymbol{u} \qquad (1)$$

$$\nabla \cdot \boldsymbol{u} = 0 \qquad (2)$$

式中：$\boldsymbol{u}$和p分别表示无量纲速度、压力；雷诺数 $Re = UC/\nu$，ν为流体的运动黏度，U 为入口处的平均速度，本研究所有算例设置为 Re=3 000。

控制方程通过有限体积法求解。对流项采用有限高斯线性方案离散化，黏性项则采用二阶中心差分方案处理。时间积分采用二阶隐式 Crank-Nicolson 格式，时间步长 $\Delta t = 5\times10^{-4}$。入口处施加自由来流边界条件，入口平均速度$U$设置为 3 m/s，出口设置为零速度梯度，并保持恒定压力；上下壁面及翼型表面设置为绝热、零压力梯度和无滑移条件。通过上述数值方法，模拟了如图 1 所示的二维计算域中翼型周围的黏性不可压缩流动。通过网格无关性验证，并在图 2 中比较了不同网格尺度下，200t^*（t^*为特征时间，即本研究的无量纲时间，定义为$t^* = t/T$，T为压力波动周期）内阻力系数的变化。3 组网格阻力系数之间误差小于 1.2%，证明计算满足网格无关性要求，能确保数值模拟的准确性和稳定性。

2 深度强化学习算法和数学模型

2.1 深度强化学习框架

DRL 作为一种闭环控制方法，通过训练神经网络获取最优控制策略，用以调节射流速度，改善流动分离带来的不利影响。

如图 3 所示的 DRL 框架，其基本可以概括为两部分和三要素。两部分指的是环境和智能体。

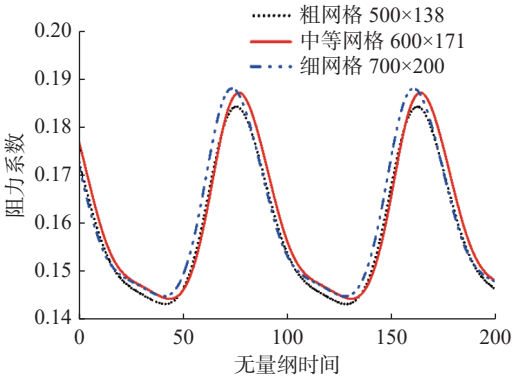


图 2　不同网格尺度下 200t^* 内的阻力系数变化

Fig.2　Drag coefficient variation during 200t^* with different grid scales

智能体通过与环境交互不断学习优化策略，以在特定任务中实现最优的控制目标。三要素指状态s_t、动作a_t、奖励r_t。状态s_t为智能体所能感知的信息合集，在本研究中表示探针探测到的流场中的速度数据；动作a_t为智能体可以执行的操作集合，本研究中表示翼型射流控制任务中的射流流量的调整；奖励函数为智能体采取某一动作的好坏，本研究使用两种奖励函数，会在后文详细解释。智能体通过接收到的奖励信息，调整其策略，从而在下一次决策时选择更好的动作。智能体从开始状态到终止状态的一个完整过程称之为回合。通过回合的进行，智能体逐步学习如何调整喷射流量，最终达到当下算例最大提升气动性能的效果。

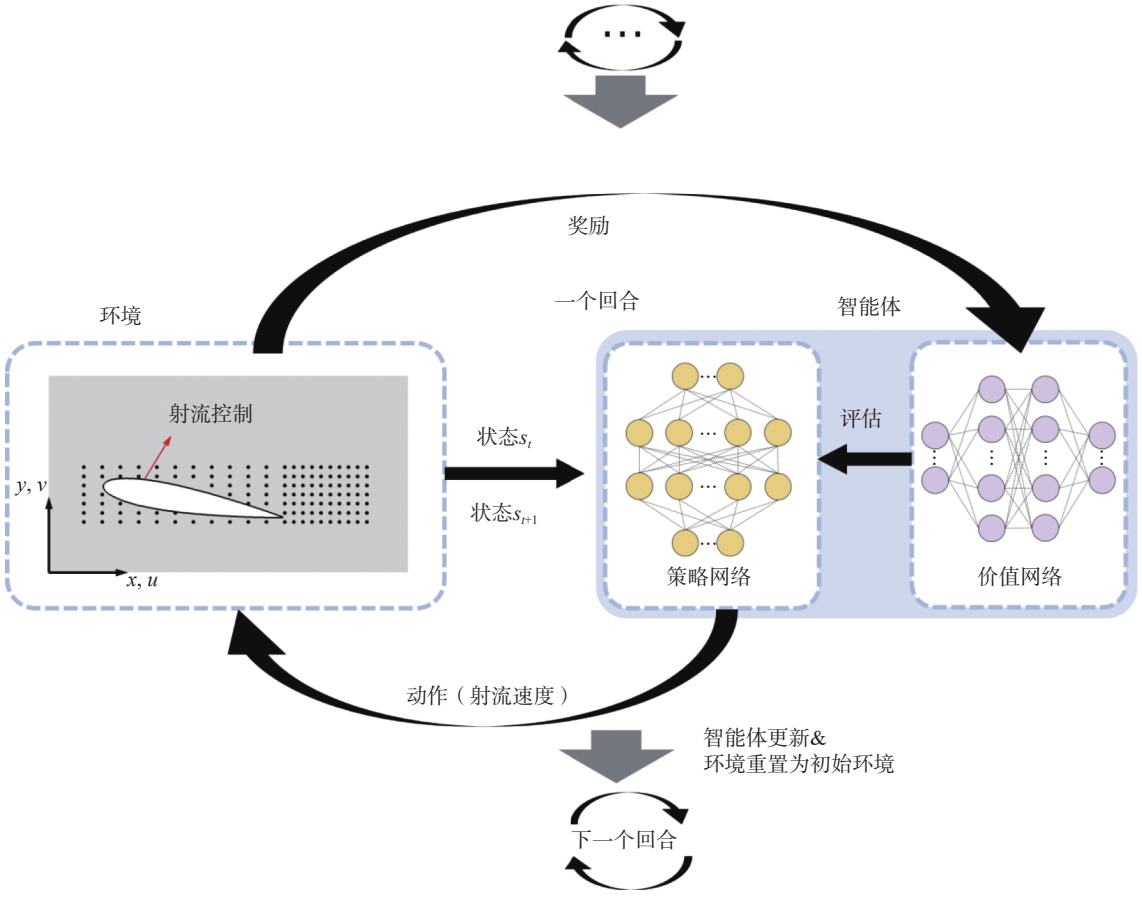


图 3　DRL 框架中射流主动流动控制的学习过程

Fig.3　Learning process of jet active flow control in DRL framework

该智能体使用 SAC 算法进行训练。策略函数和值函数由两个独立的神经网络表示，分别称为 actor（演员）和 critic（评论家），它们相互作用以评估和改进策略。actor 通过神经网络近似策略函数π，根据当前状态s_t选择动作a_t，以最大化奖励r_t。而 critic 则通过计算动作值函数$Q(s_t,a_t)$来评估策略，该函数同时使用动作a_t和当前状态s_t，并训练另一个神经网络以改进策略$\pi(a_{t+1}|s_{t+1})$。与传统的 actor-critic 方法不同，SAC 在最大熵强化学习框架中通过同时最大化奖励和熵来增强探索能力和稳定性，即使在样本较少的情况下也能有效地进行学习[17]。本研究中 actor 的人工神经网络（artificial

neural network, ANN）由两层 512×256 的全连接神经元组成。折扣因子 γ 设置为 0.97，目的是确保智能体不仅关注即时奖励，还能合理预测和利用未来的奖励，从而确保训练过程的稳定性和学习效果。软更新参数 τ 设置为 0.005，旨在平衡目标网络更新的稳定性和学习效率。采用这种软更新策略可以使目标网络的更新更加平滑，防止其过度依赖在线网络的当前状态，减少因目标网络快速变化而导致的不稳定性，并确保在 SAC 训练过程中策略网络和 Q 值网络的稳定收敛。信息熵系数 H 设置为 0.2，有助于在探索和利用之间找到良好的平衡，防止过度探索导致学习效率低下，同时避免过早收敛到次优解。学习率 β 设置为 0.001，确保 SAC 算法的稳定性和效率，并帮助算法平稳收敛到更好的策略。缓冲区大小设置为 1×10^5，保证 SAC 算法能够在复杂环境中稳定训练，并提高模型的性能和泛化能力。此外，本研究采用了 5 个并行环境进行训练，且具有相同的初始状态，每个环境从完全发展的流场开始。演员的人工神经网络由两层 512×256 的全连接神经元组成。该层的输入是 critic 网络给出的动作值函数 $Q(s_t,a_t)$，状态 s_t 则由探针（见图 1）获取的场信息传递。输出为提供射流速度的动作。critic 网络包括两组具有不同架构的神经网络，其中一组近似 Q-critic 函数，另一组近似 V-critic 函数。Q-critic 网络的输入是 s_t、a_t 和 $V(s_t)$ 的组合，输出参与更新演员网络 $Q(s_t,a_t)$。然而，V-critic 网络的输入是环境状态 H_t 和价值函数 $V(H_{t-1},H)$，包括熵 H。下一个时刻的 $V(s_t,H)$ 作为输出更新 Q-critic 网络和 V-critic 网络。

2.2 涡驱动的奖励函数

DRL 智能体在 AFC 中的一个关键作用是设计合适的奖励函数，这是因为奖励函数来自环境的直接反馈，且影响优化目标。以往关于 AFC 问题的大多数研究，尝试使用阻力和升力系数作为奖励函数。通过评估智能体在不同决策下产生的升力和阻力系数，分配相应的奖励或惩罚，从而鼓励智能体作出改善升力和减少阻力的决策。因此，首先应用一种传统的奖励函数 R_c 来控制机翼上的流动分离，其表达式如下：

$$R_c=(C_d)_{t_0}-(C_d)_t-a\left|(C_l)_{t_0}-(C_l)_t\right| \tag{3}$$

式中：$(\cdot)_{t_0}$ 表示训练前初始时刻的参数；$(\cdot)_t$ 表示每次训练后的参数；a 为常数，本研究设定为 0.05；阻力系数为 $C_d=2F_d/\rho U^2D$；升力系数为 $C_l=2F_l/\rho U^2D$。在每个时间步长中，对阻力 F_d 和升力 F_l 分别在机翼表面上进行积分，表达式为

$$\begin{cases}F_d=\int(\sigma\cdot\boldsymbol{n})\cdot\boldsymbol{e}_x\mathrm{d}S\\ F_l=\int(\sigma\cdot\boldsymbol{n})\cdot\boldsymbol{e}_y\mathrm{d}S\end{cases} \tag{4}$$

式中：$\boldsymbol{e}_x$、$\boldsymbol{e}_y$ 是一组向量，分别为 $\boldsymbol{e}_x=(1,0)$，$\boldsymbol{e}_y=(0,1)$；S 表示机翼表面面积；σ 表示柯西应力张量；$\boldsymbol{n}$ 表示垂直于翼型外表面的单位法向量。

然而，仅仅关注升力和阻力系数的整体表现，无法完全涵盖流动的基本特性，尤其是对于具有多尺度涡流生成和涡流运动的流动。基于上述考虑，本研究在奖励函数中引入 Liu 等[18] 提出的 Liutex 理论，作为优化目标之一，因为它能够表征复杂流动中的刚性涡结构，并定量涡结构的旋转强度以及涡核大小。由文献 [15] 可知，Liutex 向量定义为 $\boldsymbol{R}=R\boldsymbol{r}$，表示流体的刚性旋转量是从涡量 $\boldsymbol{\omega}$ 中分解而得到的。单位向量 $\boldsymbol{r}$ 定义为局部速度梯度张量的实特征向量，且该张量具有一个实特征值 λ_r 及一对共轭复特征值 $\lambda_{cr}\pm\lambda_{ci}$。涡旋转强度大小为 $R=\boldsymbol{\omega}\cdot\boldsymbol{r}-\sqrt{(\boldsymbol{\omega}\cdot\boldsymbol{r})^2-4\lambda_{ci}^2}$[19]。因此，考虑到为抑制大涡的生成及涡脱落的不稳定性，本研究引入了基于旋转强度和涡旋核心大小的奖励函数 R_v，其表达式如下：

$$\begin{aligned}R_v=&(C_d)_{t_0}-(C_d)_t-e\left|(C_l)_{t_0}-(C_l)_t\right|+\\ &f\Delta|R|_{\max_{\Delta t}}+g\Delta|R|_{\mathrm{mean}_{\Delta t}}\end{aligned} \tag{5}$$

式中：$\Delta|R|_{\max_{\Delta t}}$ 和 $\Delta|R|_{\mathrm{mean}_{\Delta t}}$ 分别表示从探测器获得的 Liutex 强度 $\boldsymbol{R}$ 在当前时刻的最大值与无控制状态下最大值的差值，以及当前时刻的均值与无控制状态下的均值的差值，它们的权重系数 f 和 g 分别为 0.005 和 0.05。由式（5）可知，Liutex 值的增加会导致翼型周围流场的稳定性变差。因此，$|R|_{\max}$ 和 $|R|_{\mathrm{mean}}$ 为奖励函数的负贡献。在这种情况下，考虑到阻力、升力以及当前工作中脱落涡旋的强度和核心大小，针对每一项设计了适当的权重系数。

传统的涡识别方法易受剪切层影响，而 Liutex 通过定义局部旋转强度，克服了这一物理局限，极大地提升了对复杂动态流体行为的识别精度。在主动流动控制领域，Liutex 不仅能实现涡旋结构的实时追踪，而且为操控涡动力学过程，进而优化流体系统性能提供了精确的判据。

3　结果与分析

3.1　基于 DRL 的流动分离控制

为减少阻力并稳定升力波动，针对攻角为 10°的翼型周围的分离流动，采用 DRL，并将式(3)中的奖励函数作为训练目标。训练的起始点选择在流场达到完全发展的状态，并在没有任何控制情况下形成稳定的涡街。最后，分析 DRL，进行效率和效果方面的比较。

图 4 显示了 DRL 在算例 2 中的训练过程奖励分布，式(3)中的 $(C_d)_{t_0}$ 和 $(C_l)_{t_0}$ 分别设置为 0.16 和 0.5。由图可见，在攻角 α=10°的流动条件下，两种奖励都表现出增长的趋势。DRL 在训练 240 次后呈现收敛趋势。

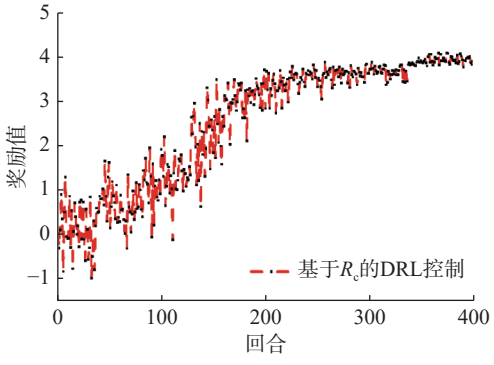


图 4　DRL 训练后算例 2 在 α=10°时的奖励值

Fig.4　Reward values for case 2 after DRL training at α=10°

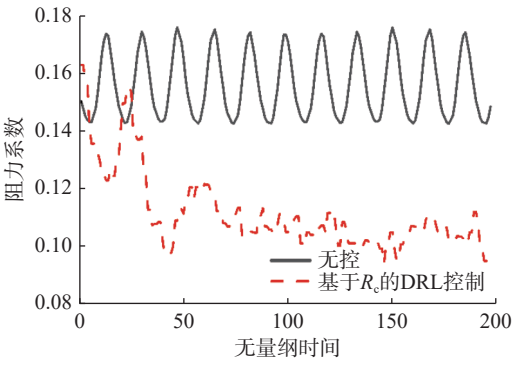
为了进一步比较和验证训练效果，图 5 比较了在 DRL 控制下，翼型周围的阻力系数 C_d 和升力系数 C_l 随时间的变化，并与无控制的情况进行对比。如图 5(a) 所示，DRL 算法显著地减少了 C_d。同时，从图 5(b) 可以看出，DRL 算法对于升力的提升也有很好的效果。算例 2 实现了 28.8% 的阻力降低和 23.7% 的升力增加。

总体而言，采用式(3)奖励函数的 DRL 智能体，一定程度上可以实现减阻和抑制升力波动，且表现出良好的性能。

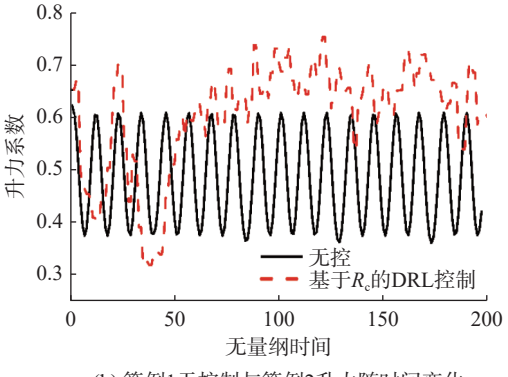
3.2　基于涡驱动的奖励机制

如上所述，分离涡直接影响机翼的空气动力学、水动力学性能。同时，控制分离的最典型特征是脱落涡的旋转强度和涡核大小。在这方面，提出了一个新的奖励函数 R_v（见式(5)），并在算例 3 中进行了验证，详细的算例描述参见表 1。

$|R|_{max}$ 和 $|R|_{mean}$，分别表示 R 的最大值和平均值，被作为式(5)中的两个新项。权重系数的确定



(a) 算例1无控制与算例2阻力随时间变化



(b) 算例1无控制与算例2升力随时间变化

图 5　DRL 控制下算例 1 和算例 2 在 α=10°时的气动参数对比

Fig.5　Comparison of aerodynamic parameters between case 1 and case 2 under DRL control at α=10°

取决于奖励函数中其他项接近 0 的程度。当 R 的值增大时，意味着涡旋变得更加活跃，流动变得更加不规则，机翼后方尾迹涡旋的稳定性下降变得更加显著。

DRL 算法在攻角为 10° 的翼型上使用新的奖励函数 R_v 进行训练(算例 3)，并与使用奖励函数 R_c 的算例 3 结果进行比较。如图 6 所示，使用 R_v 的 DRL 算法获得了更高的奖励，并且在探索阶段(0～100 轮次)有更高的增长率。如图 7(a) 所示，尽管阻力值没有显著改善，但波动的平滑效果明显增强。升力不仅有明显提升，而且与之前的结果相比，波动也经历了一定程度的抑制。

进一步地，图 8 展示了算例 3 在 3 个时刻的 Liutex 强度 R 等高线。$t_0 = 2.2T$ 对应的是接近流场初始时刻的时间点；$t_1 = 4.2T$ 时，优化策略开始影响流场，增强了涡旋的旋转强度；$t_2 = 5T$ 标志着优化策略使流场趋于稳定的时刻。可以明显看出，翼型尾缘附近脱落涡旋的旋转强度和涡核大小显著减小，这得益于应用于算例 3 的新型奖励 R_v。在图 8(c) 中的 t_3 时刻，几乎没有大的涡核结

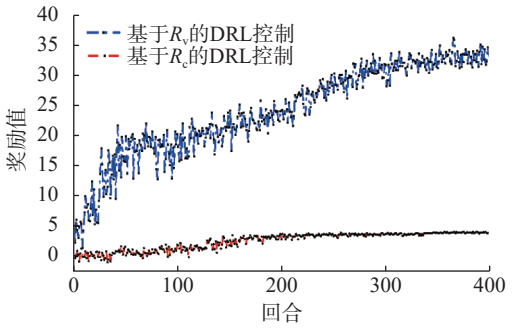
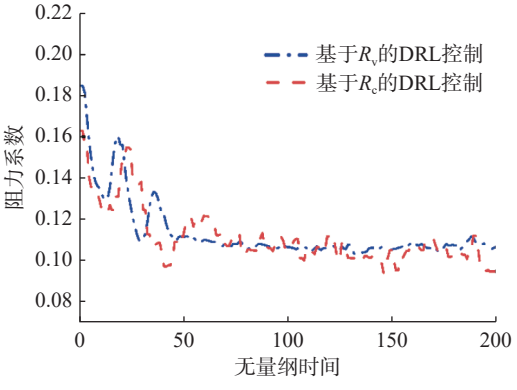
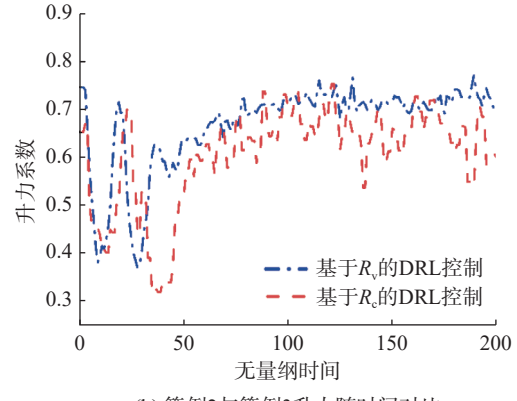


图 6 DRL 训练后算例 2 和算例 3 在 α=10°时的奖励值对比

Fig.6 Comparison of reward values between case 2 and case 3 after DRL training at α = 10°



(a) 算例2无控制与算例3阻力随时间对比



(b) 算例2与算例3升力随时间对比

图 7 DRL 控制下算例 2 和算例 3 在 α=10°时的气动性能对比

Fig.7 Comparison of aerodynamic parameters between case 2 and case 3 under DRL control at α=10°

构，也没有旋转强度的集中区域。

4 结 论

本研究采用深度强化学习方法对弱湍流条件下的翼型流动分离开展了优化控制研究。

由于传统的奖励函数在流动控制效果和控制效率方面的限制，本研究采用 Liutex 理论引入了一种新型的基于涡驱动的奖励函数 R_v，将探针接

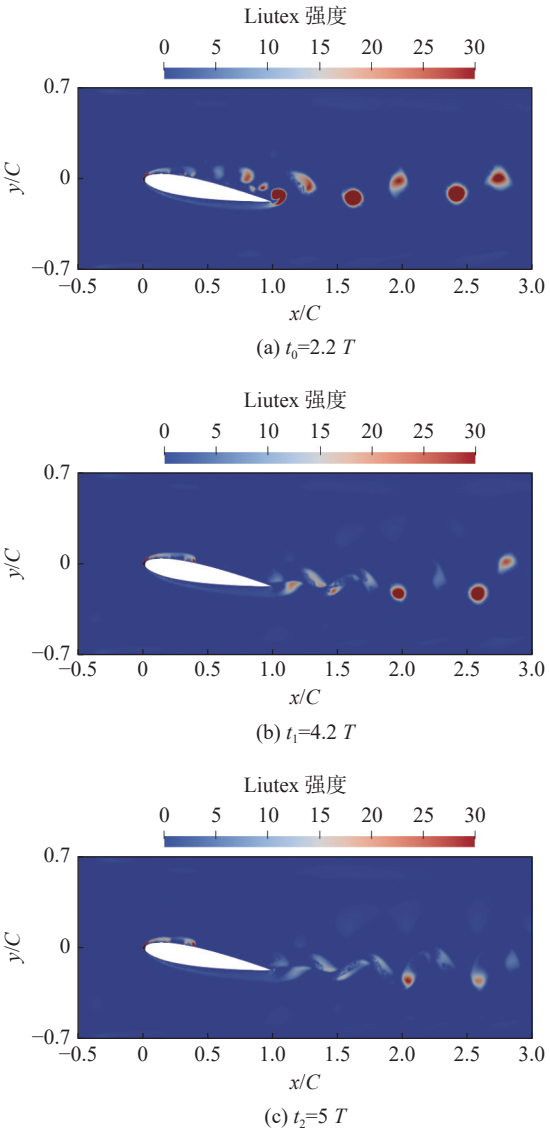


图 8 DRL 训练后算例 3 在 α=10°时的 Liutex 分布云图

Fig.8 Contours of Liutex distribution for case 2 after DRL training at α = 10°

收到的 Liutex 统计量作为奖励函数中的惩罚项，改善了机翼表面涡流的稳定性。相比于传统的涡量或 Q 标准，Liutex 能够更加精确地表征流场中的涡结构特征，特别是对刚性旋转强度及涡核区域的刻画更具物理意义。通过在奖励函数中引入 Liutex 统计量，能够有效减少机翼周围区域的 Liutex 强度，从而降低涡流的刚性旋转程度，抑制不稳定的涡核生成。这一优化策略不仅改善了机翼表面流动的稳定性，还能够在一定程度上减少流场分离，提高气动效率。数值模拟结果表明，基于 Liutex 的奖励函数在大涡抑制和流场分离问题的空气动力学优化中展现出了极大的潜力。本研究为二维翼型，未来可将该方法拓展至三维复

杂流场控制，从而进一步推广应用于高性能飞行器的流动控制、风能利用优化以及湍流减阻等领域，为航空航天与流体工程提供更高效的智能优化方案。

参考文献：

[1] JIA W, XU H. Robust and adaptive deep reinforcement learning for enhancing flow control around a square cylinder with varying Reynolds numbers[J]. Physics of Fluids, 2024, 36(5): 054103.

[2] DURIEZ T, BRUNTON S L, NOACK B R. Machine learning control-taming nonlinear dynamics and turbulence[M]. Cham: Springer, 2017: 49–68.

[3] VINUESA R, LEHMKUHL O, LOZANO-DURÁN A, et al. Flow control in wings and discovery of novel approaches via deep reinforcement learning[J]. Fluids, 2022, 7(2): 62.

[4] RABAULT J, KUCHTA M, JENSEN A, et al. Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control[J]. Journal of Fluid Mechanics, 2019, 865: 281−302.

[5] HU W L, JIANG Z Z, XU M Y, et al. Efficient deep reinforcement learning strategies for active flow control based on physics-informed neural networks[J]. Physics of Fluids, 2024, 36(7): 074112.

[6] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[C]// Proceedings of the 4th International Conference on Learning Representations. San Juan: ICLR, 2016.

[7] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[DB/OL]. (2017-08-28). https://arxiv.org/abs/1707.06347.

[8] WANG Q L, YAN L, HU G, et al. DRLinFluids: an open-source Python platform of coupling deep reinforcement learning and OpenFOAM[J]. Physics of Fluids, 2022, 34(8): 081801.

[9] DONG X R, HAO C Y, DONG Y L, et al. Investigation of vortex motion mechanism of synthetic jet in a cross flow[J]. AIP Advances, 2022, 12(3): 035045.

[10] CHOI K S, CLAYTON B R. The mechanism of turbulent drag reduction with wall oscillation[J]. International Journal of Heat and Fluid Flow, 2001, 22(1): 1−9.

[11] VIQUERAT J, RABAULT J, KUHNLE A, et al. Direct shape optimization through deep reinforcement learning[J]. Journal of Computational Physics, 2021, 428: 110080.

[12] PARIS R, BENEDDINE S, DANDOIS J. Robust flow control and optimal sensor placement using deep reinforcement learning[J]. Journal of Fluid Mechanics, 2021, 913: A25.

[13] WANG Y Z, MEI Y F, AUBRY N, et al. Deep reinforcement learning based synthetic jet control on disturbed flow over airfoil[J]. Physics of Fluids, 2022, 34(3): 033606.

[14] GONG T C, WANG Y, ZHAO X. Active control of transonic airfoil flutter using synthetic jets through deep reinforcement learning[J]. Physics of Fluids, 2024, 36(10): 106141.

[15] GUASTONI L, RABAULT J, SCHLATTER P, et al. Deep reinforcement learning for turbulent drag reduction in channel flows[J]. The European Physical Journal E, 2023, 46(4): 27.

[16] LI J C, ZHANG M Q. Reinforcement-learning-based control of confined cylinder wakes with stability analyses[J]. Journal of Fluid Mechanics, 2022, 932: A44.

[17] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018: 1861–1870.

[18] LIU C Q, GAO Y S, TIAN S L, et al. Rortex—a new vortex vector definition and vorticity tensor and vector decompositions[J]. Physics of Fluids, 2018, 30(3): 035103.

[19] WANG Y Q, GAO Y S, LIU J M, et al. Explicit formula for the Liutex vector and physical meaning of vorticity based on the Liutex-shear decomposition[J]. Journal of Hydrodynamics, 2019, 31(3): 464−474.

（编辑：丁红艺）