

# 基于强化学习的仿生机器鱼节能研究

付如彬<sup>1,2</sup> 李亮<sup>1,†</sup> 徐成<sup>2</sup> 谢广明<sup>1</sup>

1. 北京大学工学院智能仿生设计实验室, 北京 100871; 2. 92453 部队, 福州 350000;

† 通信作者, E-mail: liatli@pku.edu.cn

**摘要** 利用仿生机器鱼平台, 模拟鱼类集群游动过程, 通过实验方法探究机器鱼集群游动时的能量节省问题。具体地, 将两条机器鱼按照一定的前后距离和左右距离固定在低湍流水洞中, 并正对来流方向; 控制前方(领航)机器鱼游动姿态的变化, 使得后方(跟随)机器鱼处在变化的流场环境中。基于强化学习理论, 建立跟随机器鱼在该流场环境下游动所消耗的能量值与其游动姿态之间的映射关系; 利用该映射关系, 获得跟随机器鱼能量消耗最小值对应的游动姿态信息, 并将该信息反馈至机器鱼运动控制器中, 调整机器鱼游动姿态, 达到节省机器鱼能量的目标。

**关键词** 仿生机器鱼; 强化学习; 游动姿态; 能量节省

## Studies on Energy Saving of Robot Fish Based on Reinforcement Learning

FU Rubin<sup>1,2</sup>, LI Liang<sup>1,†</sup>, XU Cheng<sup>2</sup>, XIE Guangming<sup>1</sup>

1. Intelligent Biomimetic Design Lab, College of Engineering, Peking University, Beijing 100871; 2. 92453 PLA Troops, Fuzhou 350000; † Corresponding author, E-mail: liatli@pku.edu.cn

**Abstract** The authors studied energy saving in fish school utilizing a bio-inspired robotic fish platform. Specifically, two robotic fish were fixed in the flow tank against the current with constant left-right distance and front-back distance; the body posture of the front robotic fish was controlled to generate dynamic fluid environments, which would affect the rear robotic fish. Based on reinforcement learning, a relationship was established between energy cost and the body posture of the robotic fish. According to this relationship, the body posture could be acquired which resulted in minimum energy cost for the rear robot. This information was further transferred to the locomotion controller of the robot to adjust its body posture to save most energy.

**Key words** robotic fish; reinforcement learning; body posture; energy saving

鱼类经过亿万年的演化, 拥有非凡的水下运动能力<sup>[1-2]</sup>。即使面临复杂的外界涡流, 鱼体也能够根据外界流场的变化情况调整游动姿态, 以减小受该流场的影响, 该游动姿态称为卡门步态(Kármán gait)<sup>[3-4]</sup>。Liao等<sup>[5]</sup>探究虹鳟鱼在卡门涡街中游动时的姿态和肌肉活动情况, 发现鱼体可以通过卡门步态节省游动所消耗的能量。

为了进一步探寻鱼体如何感知外界流场的变化, 调整游动姿态, 最终实现特定的卡门步态, 节省游动中消耗的能量, 生物学家根据实验观察等方法, 提出一系列猜想: 鱼体通过侧线、内耳、鳍肢

等器官感知外界流场环境下的流速、流场压力和流场方向等信息, 并反馈到控制鱼体的中枢模式发生器(central pattern generator, CPG)中, 从而实时调整游动姿态, 实现游动时能量节省<sup>[6-7]</sup>。

受生物研究的启发, 本文基于北京大学研制的高仿生机器鱼模型<sup>[8-9]</sup>, 利用强化学习原理, 通过实验方法探究仿生机器鱼的节能问题。强化学习强调智能体如何基于环境采取行动, 从而获得最大化利益<sup>[10-12]</sup>。一般情况下, 如果没有训练数据指示, 智能体会尝试各个动作, 从结果中找到利益最大化的动作。

机器鱼在同一种流场环境下以相同的姿态游动,其能量消耗值上下波动,依据能量消耗做出的奖励行动也存在波动。所以,仅通过一次尝试,不能获得实际的奖赏值。为了避免使用遍历的方法寻找游动时能量消耗最小的动作,实验中利用强化学习原理,将机器鱼在变化流场环境中游动时的能量消耗值作为其对应环境的一种状态,建立奖赏机制,并将该奖赏机制与机器鱼的游动姿态建立相互联系,使机器鱼感知流场变化,调整为能量消耗最小的游动姿态。

本实验固定两条仿生机器鱼之间的间隔相同,且游动方向相同<sup>[13]</sup>,模拟鱼类的集群游动过程。控制前方(领航)机器鱼游动姿态的变化,使后方(跟随)机器鱼处在变化的流场环境中。基于强化学习理论,建立跟随机器鱼在该流场环境下游动所消耗的能量值与机器鱼游动姿态之间的映射关系。利用该映射关系,获得机器鱼能量消耗最小值对应的游动姿态信息,并将该信息反馈至机器鱼运动控制器中,调整跟随机器鱼游动姿态,实现游动能耗的节省。

## 1 仿生机器鱼的运动控制

### 1.1 仿生机器鱼外形结构

本文以一款仿鲇科三关节机器鱼为研究对象,如图 1 所示。该机器鱼主要由刚性头部、柔性后部鱼体和硬度递减的鱼尾组成。仿生机器鱼的主要参数如表 1 所示。

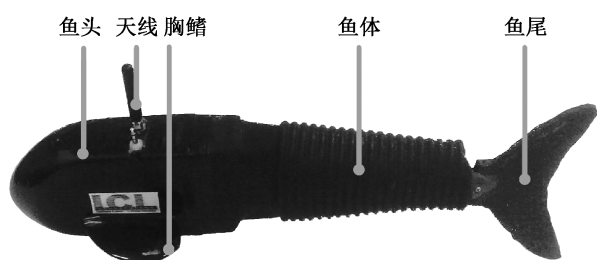


图 1 机器鱼实物照片

Fig. 1 A snapshot of our robotic fish

表 1 仿生机器鱼主要参数

Table 1 Main parameters of the robotic fish

参数	描述
尺寸(长×宽×高)	400 mm × 45 mm × 80 mm
总重量	0.8 kg
主控芯片	AVRmega128
舵机	Hitec-HS-5646WP
尾鳍展弦比	0.97

机器鱼通过控制 3 个关节的摆动形成鱼体波,边界层在水中脱落形成反卡门涡街,从而获得向前的推进力。通过染色液流动显示实验<sup>[14]</sup>,分析机器鱼游动产生的尾涡。利用高速相机从水洞底部记录机器鱼摆动时尾鳍脱落的涡街形态,观察尾涡的形成与演变过程,如图 2 所示。

图 2(a)~(d)代表 4 个比较特殊的瞬间:通过中间平衡位置时刻、即将到达最大位移时刻、离开最大位移时刻和到达中间平衡位置时刻。4 个瞬间展示一个反卡门涡对的完整形成过程。从图 2 可以看到,当尾鳍到达中间平衡位置时,脱落一个逆时针旋转的小涡,随着尾鳍向最大位移摆动,这个小涡不断变大,当尾鳍离开最大位移时,尾鳍的末端又脱落一个顺时针旋转的小涡,形成反卡门涡街。

### 1.2 包含体相位差的中枢模式发生器

CPG 是脊椎动物和非脊椎动物运动神经回路的基本组成部分,基于 CPG 模型的运动控制器广泛应用于各种仿生机器鱼的运动控制<sup>[9]</sup>。本文引入机器鱼体相位差参数  $\phi$ ,调节机器鱼体相位差参数实现游动姿态的变化,从而将外界流场信息与机器鱼摆动体相位相连接。

机器鱼摆动体相位是鱼体周期摆动身体后部和尾部时,鱼体横向偏离游动方向的偏转程度。定义机器鱼游动姿态如下:在  $t = 0$  时刻,机器鱼摆动体相位  $\phi$  为 0;一个周期后,即  $t = T$  时刻,机器鱼摆动体相位  $\phi$  为  $2\pi$ 。在摆动过程中,机器鱼的摆动体相位单调递增,如图 3 所示。

机器鱼通过感知外界流场环境,并将体相位差信号反馈至体相位 CPG 中,根据流场变化实时调整游动姿态。体相位差 CPG 控制器模型表述为

$$\dot{r}_i(t) = \alpha(R_i - r_i), \quad (1)$$

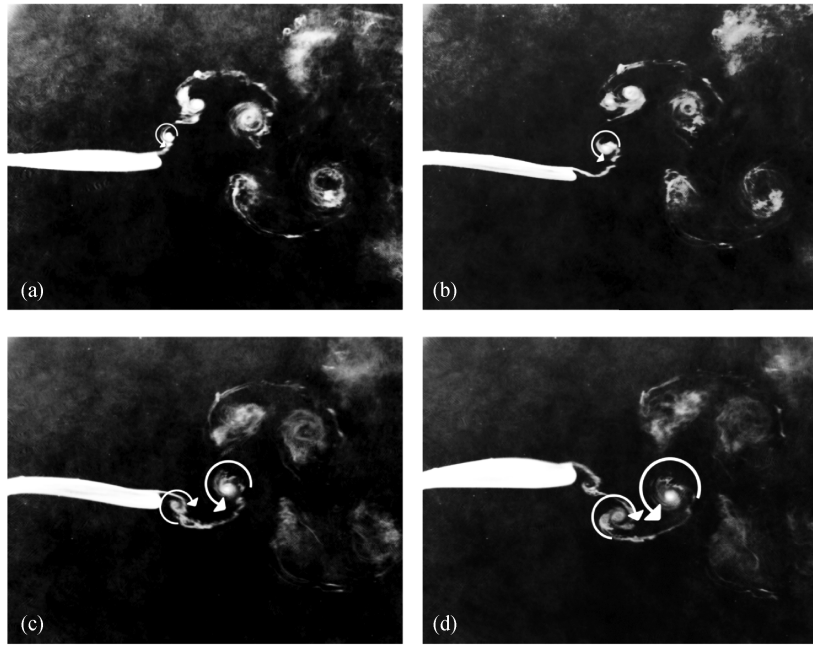
$$\dot{x}_i(t) = \beta(X_i - x_i), \quad (2)$$

$$\dot{\phi}(t) = \gamma(\Psi - \phi_i), \quad (3)$$

$$\ddot{\phi}_i(t) = \sum_{j=1, j \neq i}^N \mu[\mu(\phi_j(t) - \phi_i(t) - \phi_{ij}) - 2(\dot{\phi}_i(t) - 2\pi\omega)], \quad (4)$$

$$\theta_i(t) = x_i(t) + r_i(t)\cos(\phi_i(t) + \Phi_i). \quad (5)$$

$R_i$  和  $X_i$  表示机器鱼第  $i$  个关节的摆动幅值以及偏置角。 $r_i$ ,  $x_i$  和  $\phi_i$  表示第  $i$  个关节的幅值、偏置以及关节相位。 $\Psi$  为体相位差设定值,  $\phi_i$  为机器鱼  $t$  时刻体相位差,当  $\Psi$  变化时,  $\phi_i$  渐近收敛至设定值  $\Psi$ ,



(a) 通过中间平衡位置时刻; (b) 即将到达最大位移时刻; (c) 离开最大位移时刻; (d) 到达中间平衡位置时刻

图2 染色液流动显示结果

Fig. 2 Result of dye liquor flow display

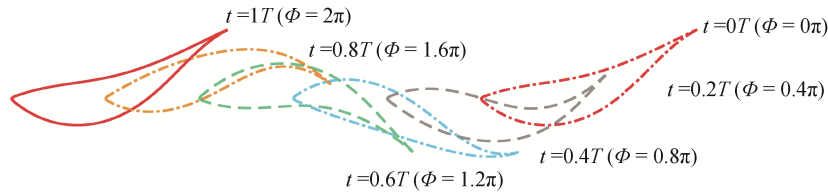


图3 机器鱼游动姿态定义

Fig. 3 Definition of the body phase

使机器鱼之间形成不同的相对游动姿态。 $\varphi_{ij}$ 表示关节*i*与*j*之间的关节相位差。 $\omega$ 为机器鱼摆动的频率。 $\alpha, \beta, \gamma, \mu$ 为结构参数,决定动态收敛的速度。 $\theta_i$ 表示第*i*个关节输出角度,由摆幅、偏置、关节相位和体相位差等参数共同决定。

实验中机器鱼的控制参数仅考虑体相位差 $\Psi$ 的变化,其他控制参数根据类似真鱼游动姿态<sup>[15]</sup>获得,见表2。

表2 机器鱼运动参数设置

Table 2 Control parameters for our robotic fish

参数	数值
关节摆幅/(°)	12, 22, 27
关节偏置角/(°)	0, 0, 0
摆动频率/Hz	0.85
关节间的相位差/rad	1.396, 2.094
结构参数	$\alpha = \beta = \gamma = 11.68, \mu = 5.84$

### 1.3 仿真验证

为了验证体相位差CPG控制器的有效性,我们进行仿真验证。两条机器鱼在同一平面,机器鱼2在前面摆动,机器鱼1在后面摆动。假设机器鱼1以相对于机器鱼2的体相位 $\Psi = 0$ 开始摆动,以相同姿态游动4 s后,给机器鱼1发送体相位差 $\Psi = \pi$ 的参数,两条机器鱼之间的相对游动姿态发生变化。仿真结果如图4所示。两条机器鱼以相同的体相位摆动4 s后,机器鱼1可以瞬间平滑地切换体相位,调整游动姿态,可以在2 s左右稳定到新的游动姿态。

## 2 基于强化学习的仿生机器鱼节能实验探究

### 2.1 实验原理

实验需要解决如何将外界流场信息与机器鱼游

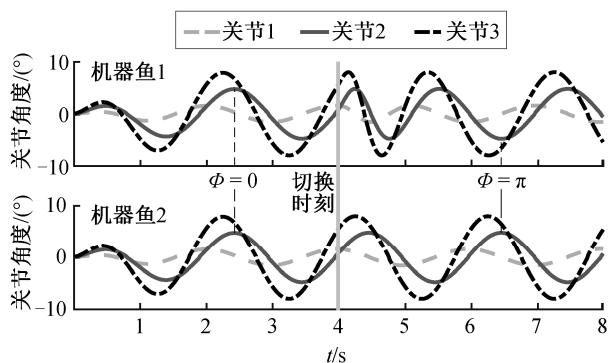


图4 仿真验证体相位差 CPG 控制器的有效性  
Fig. 4 Verification of the effectiveness of the CPG with body phase difference feedback

动姿态有效地连接,从而将机器鱼感知流场变化后的能量消耗值转换为游动姿态信息。

由于机器鱼在流场中摆动脱落的尾涡非常复杂,很难通过建立模型的方法指导机器鱼实现能量节省。本文根据强化学习策略,“探索-利用”信息,得到能量消耗最小的动作,给具有优势的动作(历史探索下能量消耗最小的动作)更多的执行机会,实现将总执行次数根据需求有倾向性地分配,从而能够更快更准地找到平均奖励最高的游动姿态,节省仿生机器鱼游动消耗的能量。

根据流场环境下不同游动姿态的能量值,采取相应策略,使机器鱼以能量消耗最小的游动姿态摆动,即建立机器鱼在流场环境下不同游动姿态与能量消耗值之间的对应关系:

$$\Psi = H(E),$$

其中,  $\Psi$  为跟随机器鱼摆动体相位差,  $H$  为强化学习算法,  $E$  为机器鱼能量消耗。基于强化学习的实验原理如图5所示。

外界流场环境由前方机器鱼通过随机切换10个离散体相位差决定,跟随机器鱼在该流场中通过强化学习获取当前能量消耗最小的体相位差。状态设置为机器鱼所处的10个流场环境,不同的流场环境对机器鱼的相互作用不同,能量消耗也有所不同,

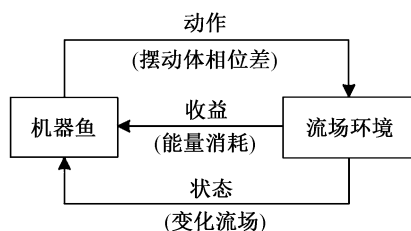


图5 基于强化学习实验示意原理图  
Fig. 5 Reinforce learning schematic diagram

动作设置为机器鱼游动姿态的10个体相位差,因而  $Q$  为  $10 \times 10$  的矩阵。实际应用中,我们可以根据不同的流场环境和游动姿态设置状态和动作,调整  $Q$  矩阵的维度。收益为机器鱼切换游动姿态后与外界流场环境交互消耗的能量。本文实验采用经典的 Q-learning 算法法则<sup>[16-17]</sup>:

$$Q(s, a) = Q(s, a) + \varepsilon(v + \eta \max_{a'} Q(s', a') - Q(s, a)),$$

其中,  $s$  表示当前机器鱼游动所处的流场环境,  $s'$  表示机器鱼游动所处的另一个流场环境,  $a$  表示机器鱼应该采取的动作(对应机器鱼能量消耗最小的体相位差  $\Psi$ ),  $a'$  表示机器鱼切换动作的游动姿态,  $v$  表示机器鱼在不同摆动体相位切换后的能量消耗,  $\varepsilon$  表示学习率,  $\eta$  表示折扣因子。

报酬函数: 本文定义强化学习的报酬函数  $v$  为机器鱼在动作  $a$  下, 摆动一段时间后消耗能量的平均值。

更新策略: 采用简单的贪婪更新策略, 即以  $(1-\epsilon)$  的概率, 采用  $Q$  值中最优动作值函数对应的体相位动作, 以  $\epsilon$  概率随机选择其他体相位动作。

## 2.2 实验设计

利用两条仿生机器鱼模拟鱼类集群游动过程, 可以通过体相位差 CPG 调整机器鱼的游动姿态, 产生不同的流场环境。相邻机器鱼感知当前流场环境, 利用强化学习, 将不同游动姿态下的能量消耗值反馈至体相位差 CPG 控制器中, 调整游动姿态, 节省机器鱼游动所消耗的能量。

控制机器鱼 A 以任意游动姿态摆动, 体相位差在  $0$  至  $2\pi$  之间变化, 间隔  $0.2\pi$ , 产生10个变化的流场环境。机器鱼 B 利用强化学习, 寻找当前流场环境中能量消耗最小的游动姿态(定义机器鱼 B 包含10个游动姿态, 每个游动姿态对应一个体相位差参数, 体相位差在  $0$  至  $2\pi$  之间变化, 间隔  $0.2\pi$ ), 将该游动姿态对应的体相位差反馈至 CPG, 调整机器鱼 B, 使其以最节省能量的体相位摆动。实验设计如图6所示。

两条机器鱼的前后、左右距离在约0.3倍体长时, 前方机器鱼摆动脱落的涡街会直接作用于跟随机器鱼, 对其能量变化产生较大的影响<sup>[18]</sup>。

将两条机器鱼固定在低湍流水洞中的同一水平面, 机器鱼 A 在前面, 机器鱼 B 在后面, 左右距离12 cm, 前后距离14 cm, 实验平台如图7所示。调

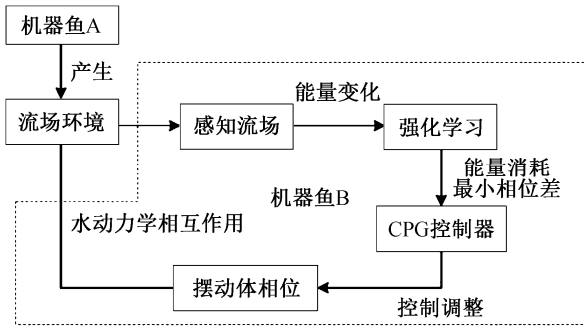


图6 实验流程设计

Fig. 6 Schematic of the experimental design

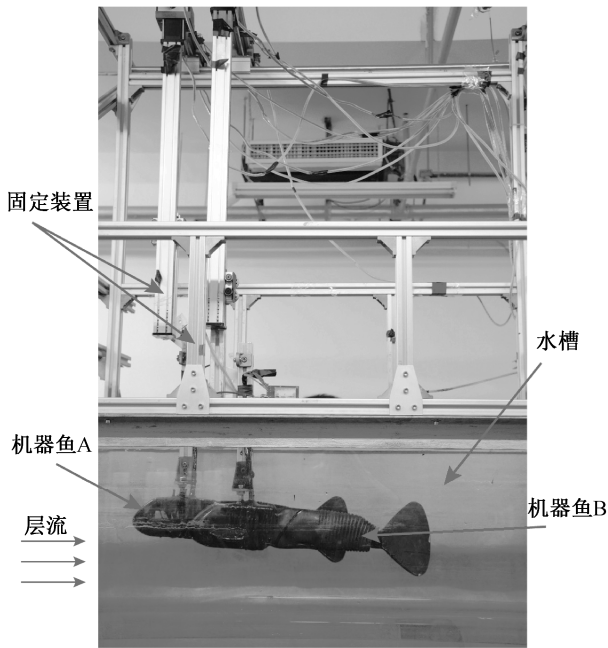


图7 实验平台

Fig. 7 Experimental platform

节水速与机器鱼自由游动速度至相近, 根据机器鱼自由游动下的游速(约 0.23 m/s), 设置水洞驱动电机的转速为 130 rpm。实验中, 能量消耗由给机器鱼供电的稳压值和通过 NI (national instruments) 电流采集卡采集的电流值获得, 采集时间为 3.4 s, 为摆动周期的整数倍(摆动频率 0.85 Hz), 确保数据的准确与稳定。

### 2.3 实验结果

第一组实验中, 基于强化学习实验原理, 采集机器鱼 B 在变化流场环境下迭代 300 次的能量消耗值。根据 Q-learning 算法, 每一个流场环境下, 机器鱼都按照能量消耗最小的策略调整游动姿态。实验中强化学习参数设置:  $\varepsilon = 0.4$ ,  $\eta = 0.8$ 。

第二组实验中, 机器鱼 B 在变化的流场环境中

自由游动, 每个流场环境下随机地给机器鱼 B 发送一个体相位差参数, 采集机器鱼 B 在变化的流场环境下的能量消耗值。两组实验结果如图 8 所示。

实验中, 能量消耗值为稳压电源的电压值乘以电流采集卡采集的电流值。定义能量效率系数  $\mu$ , 用于描述机器鱼在集群游动下的能量消耗相对于机器鱼单独游动下能量消耗的比值。

$$\mu = \frac{P_{\text{alone}} - P_{\text{school}}}{P_{\text{alone}}},$$

$P_{\text{alone}}$  表示机器鱼 B 单独游动时的平均能量消耗值,  $P_{\text{school}}$  为机器鱼 B 在集群游动时的平均能量消耗值。图 8 显示, 在集群游动时, 机器鱼 B 感知流场环境, 利用强化学习调整游动姿态游动的能量效率系数  $\mu_1 = 10.81\%$ , 机器鱼 B 随机游动的能量效率系数  $\mu_2 = 8.70\%$ , 利用强化学习实时调整游动姿态游动的能量效率比随机游动的能量效率提升约 24%。

### 3 讨论

对机器鱼的研究由来已久, 但侧重游动速度以及灵活性等方面的优化, 鲜有研究考虑机器鱼游动过程中的能量消耗问题<sup>[19]</sup>。已有的针对机器鱼游动效率优化的研究都采用间接估算的方法, 即推进力与速度的乘积。间接估算的方式仅当机器鱼以恒定速度匀速游动的时候有效, 在加速或减速的时段存在较大偏差<sup>[20]</sup>。本文直接测量功率消耗, 可以避

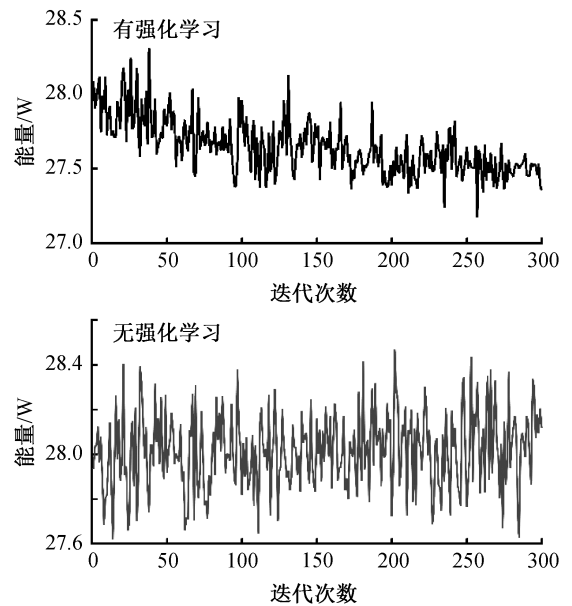


图8 有无强化学习算法下的机器鱼能量消耗对比  
Fig. 8 Comparison of energy costs of the robotic fish with and without reinforce learning

免较大偏差的出现,为后期游动效率优化的进一步研究奠定了基础。

本文引入反馈中枢神经控制器,可以将当前能耗数转换为鱼体运动的切换信号。该原理借鉴生物学上的本体感知理论,即生物体在运动过程中可以通过四肢感知外界环境对本体的施力,从而调整四肢的运动<sup>[21]</sup>。下一步可以采用信息融合的方式,引入更多感知信号,如视觉和侧线等。

本文采用强化学习的优化算法实现运动姿态的优化,从而实现能量节省。由于涡流的复杂性,基于水动力学模型的运动优化不能满足能量消耗的优化<sup>[22]</sup>,而强化学习可以避免复杂的水动力学建模,通过有监督的学习,使得机器鱼个体能够达到最优(最节省能量)的游动姿态。后期的研究可以引入无监督学习,并在更复杂的自然环境下测试本文提出的方法。

## 4 结论

本文探究了一种基于强化学习的鱼群中个体能耗优化的方法,利用强化学习,将机器鱼感知流场变化信息进行编码,并反馈到中枢神经控制器中,从而实时调整个体游动姿态,最终实现节省游动所消耗能量的目标。通过实验验证了该方法的可行性和有效性,为后续的群体编队研究奠定了基础。

将来还有很多工作需要开展。例如,基于强化学习对多条鱼进行自由编队研究,以及机器鱼的摆幅和频率对游动能耗的影响问题的研究。

## 参考文献

- [1] Webb P W. Form and function in fish swimming. *Sci Am*, 1984, 251(1): 72–82
- [2] 刘安全, 李亮, 罗文广, 等. 一种面向机器鱼的高精度位姿控制算法设计与实现. *机器人*, 2016, 38(2): 241–247
- [3] Liao J C, Beal D N, Lauder G V, et al. The Karman gait: novel body kinematics of rainbow trout swimming in a vortex street. *J Exp Biol*, 2003, 206(6): 1059–1073
- [4] Liao J C. A review of fish swimming mechanics and behaviour in altered flows. *Philosophical Transactions of the Royal Society B — Biological Sciences*, 2007, 362: 1973–1993
- [5] Liao J C, Beal D N, Lauder G V, et al. Fish exploiting vortices decrease muscle activity. *Science*, 2003, 302: 1566–1569
- [6] Mchenry M J, Michel K B, Stewart W, et al. Hydrodynamic sensing does not facilitate active drag reduction in the golden shiner (*Notemigonus crysoleucas*). *J Exp Biol*, 2010, 213(8): 1309–1319
- [7] Yanase K, Herbert N A, Montgomery J C. Disrupted flow sensing impairs hydrodynamic performance and increases the metabolic cost of swimming in the yellowtail kingfish, *Seriola lalandi*. *J Exp Biol*, 2012, 215(22): 3944–3954
- [8] Wang C, Xie G, Wang L, et al. CPG-based locomotion control of a robotic fish: using linear oscillators and reducing control parameters via PSO. *International Journal of Innovative Computing Information and Control*, 2011, 7(7B): 4237–4249
- [9] Li L, Wang C, Xie G. A general CPG network and its implementation on the microcontroller. *Neurocomputing*, 2015, 167: 299–305
- [10] 周志华. 机器学习. 北京: 清华大学出版社, 2016
- [11] Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: an introduction. *IEEE Transactions on Neural Networks*, 2005, 16(1): 285–286
- [12] Gosavi A. Reinforcement learning: a tutorial survey and recent advances. *Inform Journal on Computing*, 2009, 21(2): 178–192
- [13] Breder C. On the survival value of fish schools. *Zoologica*, 1967, 52(2): 25–40
- [14] Smits A J, Lim T T. Flow visualization, techniques and examples. London: Imperial College Press, 2000
- [15] Donley J M, Dickson K A. Swimming kinematics of juvenile kawakawa tuna (*Euthynnus affinis*) and chub mackerel (*Scomber japonicus*). *Journal of Experimental Biology*, 2000, 203: 3103–3106
- [16] Huang K Z, Yang H, King I, et al. Machine learning. *Advanced Topics in Science & Technology in China*, 2008, 22(3): 329–350
- [17] Watkins C J C H, Dayan P. Technical note: Q-learning. *Machine Learning*, 1992, 8(3/4): 279–292
- [18] Li L, Jia L, Xie G. Using robotic fish to explore the hydrodynamic mechanism of energy saving in a fish school [EB/OL]. (2015–08–13)[2018–05–10]. <https://arxiv.org/abs/1508.03249>
- [19] Yu J, Wen L, Ren Z. A survey on fabrication, control, and hydrodynamic function of biomimetic robotic fish. *Science China Technological Sciences*, 2017, 60(9): 1365–1380
- [20] Maertens A P, Triantafyllou M S, Yue D K P. Efficiency of fish propulsion. *Bioinspiration Biomimetics*, 2015, 10(4): 046013
- [21] Proske U, Gandevia S C. The proprioceptive senses: their roles in signaling body shape, body position and movement, and muscle force. *Physiol Rev*, 2012, 92(4): 1651–1697
- [22] Li L, Liu A, Wang W, et al. Bottom-level motion control for robotic fish to swim in groups: modeling and experiments. *Bioinspir Biomim*, 2019, 14: 046001