

# 控制与决策

Control and Decision

## 基于脉冲强化学习和CPG的四足机器人分层运动控制

肖云发, 韩芳, 王青云

引用本文:

肖云发, 韩芳, 王青云. 基于脉冲强化学习和CPG的四足机器人分层运动控制[J]. *控制与决策*, 2025, 40(7): 2070–2078.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2024.1029>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### [移动机器人运动规划中的深度强化学习方法](#)

Deep reinforcement learning for motion planning of mobile robots

*控制与决策*. 2021, 36(6): 1281–1292 <https://doi.org/10.13195/j.kzyjc.2020.0470>

#### [带有输出约束的柔性关节机械臂预设性能自适应控制](#)

Prescribed performance adaptive control of flexible-joint manipulators with output constraints

*控制与决策*. 2021, 36(2): 387–394 <https://doi.org/10.13195/j.kzyjc.2019.0974>

#### [基于改进卷积神经网络的动力下肢假肢运动意图识别](#)

Intent recognition of power lower-limb prosthesis based on improved convolutional neural network

*控制与决策*. 2021, 36(12): 3031–3038 <https://doi.org/10.13195/j.kzyjc.2020.0326>

#### [基于生物启发神经网络和DMPC的多机器人协同搜索算法](#)

Multi-robot cooperative search algorithm based on bio-inspired neural network and DMPC

*控制与决策*. 2021, 36(11): 2699–2706 <https://doi.org/10.13195/j.kzyjc.2020.0959>

#### [气动人工肌肉驱动的机器人控制方法研究现状概述](#)

Overview of control methods for pneumatic artificial muscle-actuated robots

*控制与决策*. 2021, 36(1): 27–41 <https://doi.org/10.13195/j.kzyjc.2020.0793>

# 基于脉冲强化学习和 CPG 的四足机器人分层运动控制

肖云发<sup>1</sup>, 韩芳<sup>1,2†</sup>, 王青云<sup>2,3</sup>

(1. 东华大学 信息科学与技术学院, 上海 201620; 2. 宁夏大学 数学统计学院, 银川 750021;  
3. 北京航空航天大学 动力学与控制系, 北京 100191)

**摘要:** 随着四足机器人控制技术的逐渐发展, 四足机器人已经广泛被应用于救援、军事、探险等领域. 在四足机器人的应用中, 如何在不牺牲控制性能的前提下有效降低能耗是一个重要的瓶颈问题. 为此, 提出一种基于脉冲强化学习算法 (SRL) 和中枢模式发生器 (CPG) 的分层控制算法 (SRL-CPG) 用于四足机器人的运动控制. 首先, 因为脉冲神经元相比传统的人工神经元具有更低的能耗, 故基于脉冲神经网络 (SNN) 构建脉冲强化学习算法, 将其作为控制中枢; 其次, 在控制任务动作空间过大的情况下, SRL 难以取得良好的控制效果, 因此将 CPG 作为低级控制器, 利用 SRL 接受状态信息并对 CPG 参数进行调整进而控制四足机器人运动; 最后, 将 SRL-CPG 控制算法在 Webots 环境中搭建的四足机器人模型 Gbot 上进行实验验证, 结果表明 SRL-CPG 控制算法能够有效应用于四足机器人的运动控制, 并大大降低能耗.

**关键词:** 脉冲神经网络; 中枢模式发生器; 四足机器人; 脉冲强化学习; 分层控制; 运动控制

中图分类号: TP242.6 文献标志码: A

DOI: 10.13195/j.kzyjc.2024.1029

引用格式: 肖云发, 韩芳, 王青云. 基于脉冲强化学习和 CPG 的四足机器人分层运动控制 [J]. 控制与决策, 2025, 40(7): 2070-2078.

## Hierarchical motion control of quadruped robot based on spiking reinforcement learning and CPG

XIAO Yun-fa<sup>1</sup>, HAN Fang<sup>1,2†</sup>, WANG Qing-yun<sup>2,3</sup>

(1. College of Information Science and Technology, Donghua University, Shanghai 201620, China; 2. School of Mathematics and Statistics, Ningxia University, Yinchuan 750021, China; 3. Department of Dynamics and Control, Beihang University, Beijing 100191, China)

**Abstract:** The control technology of quadruped robots has gradually developed, and quadruped robots have been widely applied in fields such as rescue, military, and exploration. In the application of quadruped robots, how to effectively reduce energy consumption without sacrificing control performance is an important bottleneck problem. In response to this problem, this paper proposes a hierarchical control algorithm based on the spiking reinforcement learning (SRL) algorithm and the central pattern generator (CPG). First, considering that spiking neurons have lower energy consumption than traditional artificial neurons, this paper constructs a SRL algorithm based on a spiking neural network (SNN) and uses it as the control center. Then, when the action space of the control task is too large, it is difficult for the SRL algorithm to achieve good control effects, thus the CPG model is used as the low-level controller, and the SRL algorithm is utilized to accept state information and adjust CPG parameters for controlling the movement of the quadruped robot. Finally, the SRL-CPG control algorithm is experimentally verified on the quadruped robot model Gbot built in the Webots environment. The results show that the SRL-CPG control algorithm can be effectively applied to the motion control of quadruped robots and significantly reduce energy consumption.

**Keywords:** spiking neural network; central pattern generators; quadruped robot; spiking reinforcement learning; hierarchical control; motion control

## 0 引言

随着机器人控制技术的发展, 四足机器人已广

泛应用于工业、救援、教育和娱乐等多个领域<sup>[1]</sup>. 宇树科技的 UnitreeGo 系列在工业和救援领域表现突

收稿日期: 2024-08-29; 录用日期: 2024-12-04.

基金项目: 国家自然科学基金项目 (12272092, 12332004).

责任编辑: 贾英民.

†通信作者. E-mail: yadiahhan@dhu.edu.cn.

出,云深处科技的绝影系列在科研教育发挥关键作用,而小米的铁蛋系列则在娱乐领域带来创新。然而,实现这些机器人在离线状态下的长期自主运行仍是一个技术挑战,关键在于如何在降低能耗的同时保持其控制性能,确保在多变环境中的稳定性和效率。

在四足机器人的早期研究中,模型预测控制(MPC)和基于振荡器的步态生成方法被广泛应用于运动控制<sup>[2]</sup>。然而,这些传统方法要求对四足机器人进行精细建模,对环境变化的适应性有限,难以在多变的环境中有效执行任务。近年来,随着人工智能技术的飞速发展,研究者开始将神经网络模型融入四足机器人的运动控制策略中<sup>[3]</sup>。例如,孙辉辉等<sup>[4]</sup>提出了一种基于强化学习的控制方案,通过预测机器人的期望行进位置,成功实现了在复杂地形下按指定方向的直线行走任务。在深度学习的推动下,深度神经网络(DNN)作为第2代神经网络的代表,已成为人工智能领域的主导模型<sup>[5]</sup>。但是,由于DNN的高计算需求和能耗,其在需要实时响应的仿生机器人应用中受到了限制。脉冲神经网络(SNN)作为第3代神经网络,模仿生物神经元的机制构建,以其低计算量、低功耗和高非线性度而受到关注。SNN能够通过时间、相位、频率等多维度对输入信号进行编码,并模拟生物神经元的放电过程<sup>[6]</sup>,这些特性使得SNN具备了执行复杂信息处理的能力<sup>[7]</sup>。因此,研究者开始探索将SNN应用于仿生机器人的运动控制领域。Nichols等<sup>[8]</sup>提出了一种SNN结构,使机器人能够在仿生平台上记忆环境并追随目标。Batllori等<sup>[9]</sup>开发了一种基于遗传算法优化的SNN,使机器人能够在有障碍物的环境中导航至指定位置。此外,脉冲强化学习(SRL)结合了SNN和强化学习(RL),在处理运动规划问题时具有巨大的优势<sup>[10]</sup>。也有科学家将脉冲强化学习算法引入运动控制中<sup>[11]</sup>,但是当控制任务变得复杂时,使用SRL作为单一控制层的控制策略在训练过程中可能会耗费大量时间,且模型难以收敛。

中枢模式发生器(CPG)是哺乳动物体内的一种神经元回路,能够在无外界信号输入的情况下产生节律行为,这些行为不仅易于调节,还有鲁棒性和高耦合等优点,使其成为理想的运动信号发生器,并在机器人领域得到广泛应用<sup>[12-14]</sup>。CPG模型主要分为基于神经元建模和基于非线性振荡器建模两大类<sup>[14]</sup>。Matsuoka和Kimura模型是神经元模型CPG的代表,而Kuramoto相位振荡器和Hopf谐波振荡器则是振荡器模型CPG的主流<sup>[15-17]</sup>。随着CPG在四足机器人运动控制中的广泛应用,Liu等<sup>[18]</sup>提出了一种

多级CPG模型(ML-CPG),用以实现四足机器人的自适应控制。尽管基于CPG的四足机器人运动控制算法建模简单和调控便利,但其在整合机器人状态反馈方面的能力有限,限制了其在复杂环境中的表现。

也有科学家尝试用神经网络作为控制中枢,其他控制器作为低级控制器的分层控制架构<sup>[19-20]</sup>。Bellegarda等提出一种分层地形感知控制框架,利用深度强化学习(DRL)进行高级规划,并应用最优控制作为低级控制器,完成了四足机器人的速度控制任务。实验结果表明,分层控制方案相较传统控制方案具有更好的稳定性和更优的控制效果。分层控制架构可以让每一层都可以专注于其特定的功能,层与层之间的交互通过定义良好的接口进行,使得各层之间的耦合度降低,从而提高控制器的效率和准确性<sup>[21]</sup>。但是,面向四足机器人的运动控制任务,分层控制器层之间的具体任务分工以及合作模式仍需进一步探索。

本文提出一种基于脉冲强化学习和CPG的四足机器人分层运动控制算法(SRL-CPG)。该算法应用分层控制架构设计,SRL负责接收环境反馈的状态信息,并据此动态调整低级控制器的CPG参数,由此简化了SRL的动作空间,并解决了CPG参数难以自适应环境变化的问题。在Webots环境中进行的实验结果表明:所提出的SRL-CPG方法能在满足直行和速度控制任务的需求下输出稳定和自适应能力强的动作策略,并且能够大幅度降低控制策略的能耗。该方法将脉冲强化学习引入四足机器人运动控制中,并采用新颖的分层架构设计算法,为四足机器人节能运动控制提供了一种参考策略,使得四足机器人可以在降低能耗的前提下,在多种任务中保持良好的运动性能。

## 1 问题描述和整体控制框架

### 1.1 问题描述

本文旨在开发一种节能型四足机器人运动控制算法,该算法在执行运动控制任务时,能够在不牺牲控制性能的基础上,实现能耗的显著降低。可以将四足机器人的运动控制问题转化为马尔可夫决策过程(MDP),智能体依据对应环境产生对应的行为。

MDP由4组变量( $S, A, P, R$ )组成, $S$ 表示状态, $A$ 表示智能体此时有效的动作, $P$ 表示状态转移概率, $P(s_{t+1}|s_t, a_t)$ 表示从状态 $s_t$ 采取动作 $a_t$ 转换到状态 $s_{t+1}$ 的概率, $R(s_t, a_t, s_{t+1})$ 表示从状态 $s_t$ 采取动作 $a_t$ 转换到状态 $s_{t+1}$ 的回报,智能体的目标是学习一

个策略 $\pi(s)$ ,使给定任务的预期回报最大化.

状态空间选择需要兼顾状态的可获取性和满足训练需求的必要性.本文选取16维状态信息作为状态空间,其具体构成为机身质心的 $X$ 、 $Y$ 、 $Z$ 方向实时位移 $(d_x, d_y, d_z)$ ,机身横滚、俯仰、偏航角度(roll, pitch, yaw),机身质心 $X$ 、 $Y$ 、 $Z$ 方向实时速度 $(v_x, v_y, v_z)$ ,机器人足端是否接触地面 $(g_1, g_2, g_3, g_4)$ ,控制目标 $X$ 、 $Y$ 轴方向速度 $(v_x^{targ}, v_y^{targ})$ 以及偏转角 $(yaw^{targ})$ .状态空间用16维向量表示,即

$$o_t = [s_1, s_2, \dots, s_{15}, s_{16}]. \quad (1)$$

动作空间 $\hat{a}_t \in \mathbb{R}^{12}$ 为机器人8个关节角与下一个时间步的期望位置与初始位置的偏差 $a_t$ .所有的动作信息都为连续值,用8维向量表示为

$$\hat{a}_t = [a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8]. \quad (2)$$

因此,期望关节角 $\hat{q}_t$ 等于 $\hat{q}$ 与 $a_t$ 乘上比例系数 $\sigma_a$ 之和.获取到期望关节角度后,本文采用比例微分控制器将期望角度 $\hat{q}_t$ 转化为期望力矩

$$\hat{\tau}_t = K_p(\hat{q}_t - q_t) + K_d(\hat{q}_t - \dot{q}_t), \quad (3)$$

其中 $K_p$ 和 $K_d$ 分别为比例系数和微分系数.在每一个特定的时刻,对四足机器人各个关节的力矩进行控制,就可以控制四足机器人4条腿的动作,完成四足机器人运动控制任务.

### 1.2 整体控制框架

在生物学中,动作信号经由脊髓传递至效应器,效应器完成相应的动作<sup>[21]</sup>.受此启发,本文设计了一种模拟生物运动机理的四足机器人分层运动控制框架,如图1所示.脉冲强化学习模拟大脑皮层的功能,接收融合了控制指令和机器人状态信息的种群编码输入.SRL根据脉冲输入,向中央模式发生器输出动作信号,这些信号通过饱和函数传递至CPG.CPG调整其网络结构,并向比例-微分(PD)控制器发送每个关节的期望位置.PD控制器根据这些输入计算并输出关节力矩控制信号至机器人,驱动各关节转动,实现预定动作.

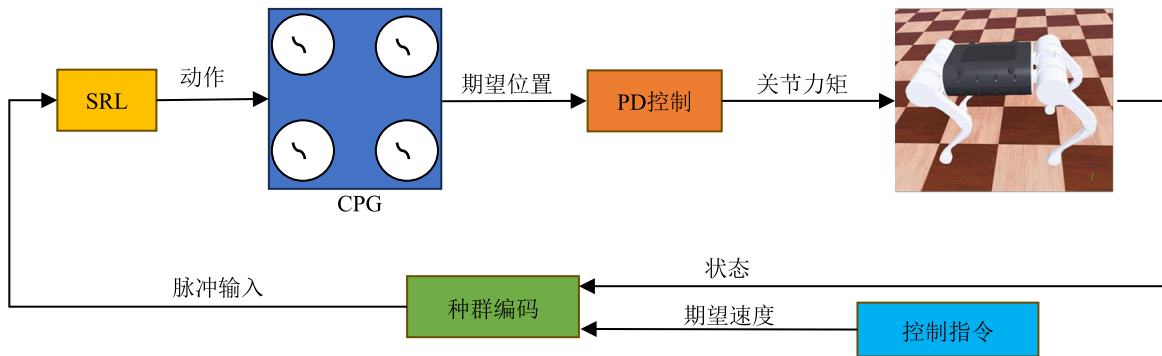


图1 基于SRL-CPG算法的四足机器人运动控制方案

## 2 SRL-CPG 分层运动控制算法

算法结构如图2所示,其输入包含给定速度指令和四足机器人的状态信息,在种群编码之后,脉冲信息输入脉冲actor网络中进行计算;脉冲actor网

络应用混合训练模式,在训练的过程中与深度critic网络共同训练,其接受脉冲输入之后,输出经过种群解码后将选择CPG的参数 $\mu, \omega$ ,并且也会直接调整CPG的输出;最后模型通过逆运动学,将CPG的

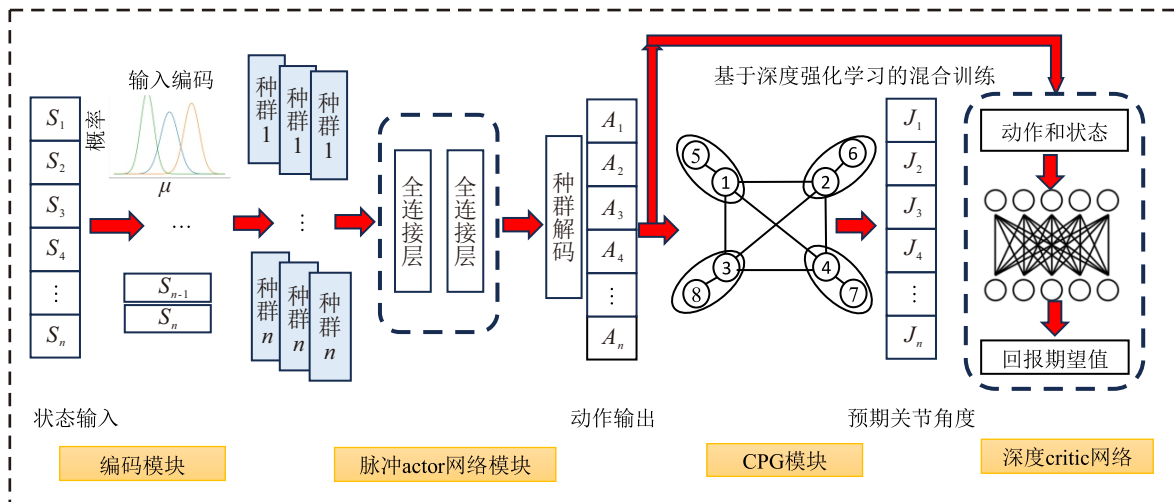


图2 SRL-CPG模型示意图

状态转换成四足机器人的预期关节角度,输入PD控制器之中。

## 2.1 脉冲神经元模型

SNN 通过脉冲序列进行信息的传输。目前,研究人员已经提出多种脉冲神经元模型,但在 SNN 框架中,LIF(leaky integrate and fire)模型是应用最广泛的模型之一。LIF神经元的膜电位动态如下:

$$H_t = f(V_{t-1}, X_t), \quad (4)$$

$$S_t = \Theta(H_t - V_{th}), \quad (5)$$

$$V_t = H_t(1 - S_t) + V_{reset}S_t. \quad (6)$$

其中:  $H_t$ 和 $V_t$ 分别表示时间步长为 $t$ 时脉冲发放前后的膜电位,  $X_t$ 表示外部输入,  $S_t$ 表示输出脉冲。式(4)~(6)描述了膜电位的发放前后的动态变化。

如果有脉冲输出,则其值为1,否则其值为0。 $V_{th}$ 为神经元的脉冲发放阈值电压,  $V_{reset}$ 为神经元的膜电位复位电压。式(4)~(6)建立了一个通用的数学模型来描述脉冲神经元的离散动态,包括充电、放电和复位3个过程。具体来说,式(4)描述了膜电位充电动态。式(5)描述了放电动态,脉冲生成函数 $\Theta(x)$ 为阶跃函数,其定义为:当 $x \geq 0$ 时,  $\Theta(x) = 1$ ;当 $x < 0$ 时,  $\Theta(x) = 0$ 。式(6)描述了膜电位复位动态,脉冲神经元采用硬复位的方式,神经元发放过脉冲后,膜电位变成复位电压。其中LIF神经元的函数 $f(\cdot)$ 定义如下:

$$f(V_{t-1}, X_t) = \alpha_V(V_{t-1} - V_{reset}) + V_{reset} + X_t. \quad (7)$$

其中: $\alpha_V$ 为电压衰减因子,当 $\alpha_V=1$ 时,式(7)即为IF神经元的函数 $f(\cdot)$ 。

## 2.2 SRL模型

本小节介绍分层架构中SRL的实现细节。在训练过程中利用深度critic网络辅助脉冲actor网络(SAN)进行训练,并且采用种群编码方式对输入编码,SAN主要分为脉冲编码模块、脉冲神经网络计算模块、脉冲解码模块。

1) 脉冲编码模块。对于 $N$ 维状态 $S \in \mathbb{R}^N$ ,应用不同高斯分布域( $\mu, \sigma$ )的神经元种群 $E$ 将每个状态维度 $s_i$ 编码成一个脉冲序列,其中 $\mu$ 和 $\sigma$ 是任务特定的可训练参数。如果 $P_{in}$ 表示每个状态维的输入总体大小,则脉冲序列的形状为 $(NP_{in}, T)$ 。首先,将状态 $S$ 转换为群体中每个神经元的刺激强度 $A_E$ ,转换过程如下:

$$A_E = e^{-\frac{1}{2} \left( \frac{s_i - \mu}{\sigma} \right)^2}. \quad (8)$$

本文使用确定性编码,将 $A_E$ 作为 $t$ 时间步是否发放脉冲的概率,本文将梯度与刺激强度 $A_E$ 反向传播,

在任何时间步长 $t$ 上如果发射脉冲,则 $\delta S_t / \delta A_E = 1$ ,否则 $\delta S_t / \delta A_E = 0$ 。

2) 脉冲神经网络计算模块。脉冲神经网络计算模块连接种群编码器和种群解码器,它们的输入和输出是脉冲序列。SNN的典型结构包括突触层和神经元层。突触层包括卷积层和全连接层,每一层后面都有一个神经元层。由于种群编码器在处理高维数据时会带来极高的计算成本,本文将状态特征提取为一维向量,并仅使用FC层作为突触层。此外,LIF神经元形成神经元层,所有权重参数在所有仿真时间步都是共享的。对于具有 $M$ 维动作的任务,本文将最后一层的脉冲神经元平均划分为 $M$ 个输出种群,大小为 $P_{out}$ 。每个输出种群都有相应的种群解码器。

3) 种群解码模块。其中种群 $D_i$ 表示 $M$ 维动作 $A \in \mathbb{R}^M$ 的第 $i$ 维。解码器模块将脉冲输入转换成连续值输出分为两个阶段。首先,在每 $T$ 个时间步之后,对 $D_i$ 中每个神经元的脉冲求和,得到在 $T$ 个时间内该种群神经元的平均放电率 $f_r$ ;然后, $f_r$ 加权求和之后输出动作 $a_i$ 。解码器输出时的全连接部分连接权重也会在网络训练的过程中被学习。

## 2.3 CPG模型

主流CPG模型按照建模方式可分为两种:基于神经元建模和基于非线性振荡器建模。前者模型虽然比较贴合动物的神经系统,但控制参数较多,模型复杂,不适合用作低级控制器。后者模型相对简单,参数易于设置和调整。考虑到Hopf振荡器模型数学描述简单,可以自行调整参数,故本文基于此搭建低级控制器模型。单个Hopf振荡器的数学表达式如下:

$$\begin{cases} \dot{x} = \alpha(\mu - r^2)x - \omega y; \\ \dot{y} = \alpha(\mu - r^2)y + \omega x. \end{cases} \quad (9)$$

其中: $x$ 和 $y$ 是振荡器的状态变量,描述振荡器的输出曲线; $\alpha$ 是一个常数,决定收敛速度( $\alpha = 1000$ ); $\mu$ 是分岔参数, $\mu > 0$ 时,振荡器才会发生振荡,决定振荡器的输出幅值( $A = \sqrt{\mu}$ ); $\omega$ 表示振荡器的震荡频率。随着 $\mu$ 和 $\omega$ 变化,振荡器输入的变化曲线如图3所示。

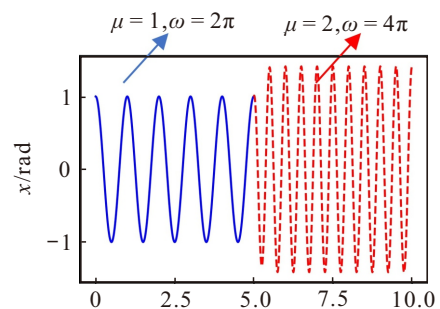


图3 振荡器输出变化曲线

本文中的 CPG 模型由 4 个 Hopf 振荡器组成, 每个振荡器都对应控制一条腿的周期性运动, 4 个振荡器之间的耦合关系如下:

$$\begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} = \begin{bmatrix} a(\mu - r_i^2) & -\omega_i \\ \omega_i & a(\mu - r_i^2) \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \sum_{j=1}^4 R(\theta_{ij}) \begin{bmatrix} x_j \\ y_j \end{bmatrix}, \quad (10)$$

其中  $R(\theta_{ij})$  是耦合矩阵, 表示 4 个振荡器之间的耦合关系. 此外, 四足机器人髋关节  $\theta_{hi}$  和膝关节  $\theta_{ki}$  具有强耦合性, 其关系如下:

$$\begin{cases} \theta_{hi} = x_i; \\ \theta_{ki} = \begin{cases} -\operatorname{sgn}(\psi) \frac{A_k}{A_h} y, & y < 0; \\ 0, & y \geq 0. \end{cases} \end{cases} \quad (11)$$

其中:  $\operatorname{sgn}(\psi)$  表示符号函数, 其函数值与  $\psi$  正负相关, 若  $\psi \geq 0$ , 则  $\operatorname{sgn}(\psi) = 1$ , 否则  $\operatorname{sgn}(\psi) = -1$ . 在  $t$  时刻, CPG 控制最终向 PD 控制器发出四足机器人前行时 8 个关节的 8 维的期望角度信号  $J$ , 进而控制运动动作的生成.

综上, 本文构建了 SRL-CPG 模型, 该模型主要是以 SRL 为高级控制器, 调整 CPG 的参数进而控制 CPG 的输出, CPG 作为低级控制器接收高级控制器的指令进而生成四足机器人运动步态. 本文采用这样的分层架构, 可以充分发挥 SRL 的自适应能力和 CPG 的步态生成能力.

### 3 四足机器人运动控制仿真实验和结果

#### 3.1 实验设置

本文选择 Webots 仿真平台对开发四足机器人 Gbot 进行仿真建模, 并对所提出的算法进行实验.

Gbot 机器人共有 4 条腿, 每条腿有膝关节和髋关节两个关节, 其中髋关节占主导作用, 膝关节随着髋关节的变化而变化. 机器人详细参数如表 1 所示.

表1 Gbot 仿真参数

序号	参数	取值
1	$M_{\text{body}}/\text{kg}$	6.5
2	$m_{i1}/\text{kg}$	0.45
3	$m_{i2}/\text{kg}$	0.35
4	$L_{\text{body}}/\text{m}$	0.27
5	$W_{\text{body}}/\text{kg}$	0.24
6	$H_{\text{body}}/\text{kg}$	0.09
7	$L_1/\text{m}$	0.25
8	$L_2/\text{m}$	0.20

机身质量为  $M_{\text{body}}$ ,  $L_{\text{body}}$ 、 $W_{\text{body}}$ 、 $H_{\text{body}}$  分别代表机身长宽高. 两个关节之间连杆长度为  $L_1$ , 质量为

$m_{i1}$ ; 关节与足端之间连杆长度为  $L_2$ , 质量为  $m_{i2}$ .

实验中使用的模拟环境用 Webots 接口, 实验地形可通过设置参数进行调控, 本文使用 PyTorch 作为深度学习框架. 模拟实验是在一台带有 Intel-i7 处理器、16 GB RAM 和 NVIDIA RTX4050 的电脑上进行. 对于直行任务, 每个实验的训练时间约为 1h. 对于速度控制任务, 每个实验的训练时间约为 3h. 为了评估所提出方法, 本文提供了与使用相同 SRL 算法训练的策略和单层 CPG 策略进行对比. 表 2 中给出了用于训练 SRL-CPG 策略超参数的详细信息.

表2 SNN-CPG 参数配置

参数	值
Learning rate	0.001
Reward discount factor	0.99
Exploration noise with stddev	0.1
Length of replay buffer	1000000
Soft target update factor	0.005
Batchsize	100
Update steps	1
Policy update delay	1
CPG Policy steps	5

#### 3.2 复杂地形下直行任务

SRL 已经被证实能在直行任务中达到深度强化学习的水平, 本文在直行实验中, 将 SRL-CPG、SRL、CPG 三者进行对比来评估本文算法相对原有算法的改进. 为使机器人保持直行, 奖励函数设计如下:

$$\begin{cases} r_t = \xi_{\text{dist}} d_x - \xi_{\text{dev}} d_y - \xi_{\gamma} |\gamma| - \xi_j s_{\text{joint}}, \\ d_x = x_{t+1} - x_t, \\ d_y = |y_{t+1}| - |y_t|. \end{cases} \quad (12)$$

其中:  $d_x$  表示机器人质心在两个时间步之间的  $X$  方向位移;  $d_y$  表示机器人的横向偏移距离;  $\gamma$  表示机器人的偏移方向;  $s_{\text{joint}}$  表示机器人行进的过程中关节之间是否出现异常动作, 对其进行惩罚. 实验中,  $\xi_{\text{dist}} = 1$ ,  $\xi_{\text{dev}} = 0.3$ ,  $\xi_{\gamma} = 0.3$ ,  $\xi_j = 0.001$ .

在 Webots 中改变地形复杂程度参数的值可以改变地形的复杂程度, 本文设计直行实验在 3 种复杂程度递增的地形 ENV1、ENV2、ENV3 中进行. 模型训练过程中 SRL-CPG 和 SRL 模型都是基于 ENV2 进行训练.

本文将 SRL 模型和 CPG 模型引入对比试验, 每间隔一个时间步, 采样机器人位置, 本文共在 100 个时间步内进行采样, 实验结果如图 4 所示. 运动过程中记录偏移距离, 如表 3 所示.

在复杂程度最低的 ENV1 环境中, 3 种模型均只出现了较小的偏离; 随着环境复杂程度上升, CPG 模型前行方向出现较大偏离, SNN 模型也在第 3 种

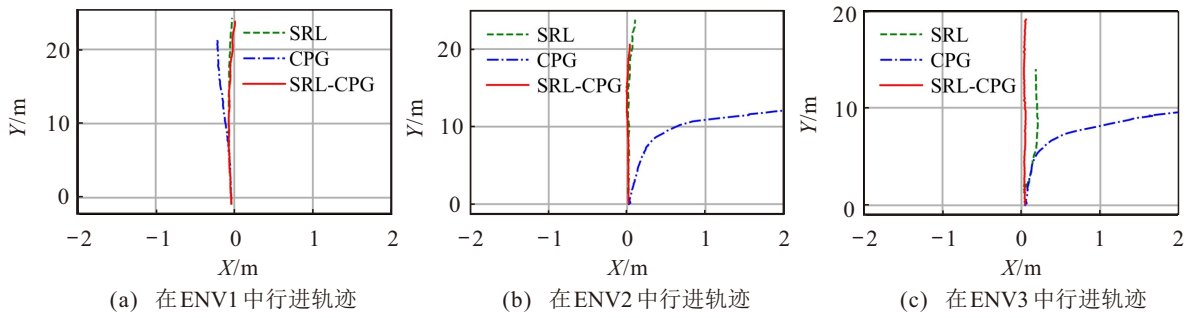


图4 3种地形下四足机器人的行进轨迹

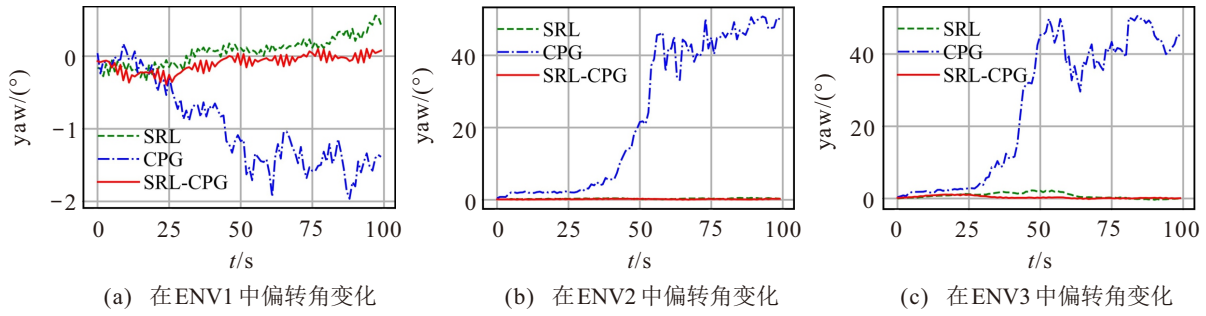


图5 3种地形下四足机器人的偏转角变化

表3 四足机器人偏移距离

环境	偏移距离/m		
	SRL	CPG	SRL-CPG
ENV1	-0.03	-0.21	0.01
ENV2	0.10	6.59	0.03
ENV3	0.18	6.91	0.05

环境中出现明显偏离, SNN-CPG 模型在出现一定偏离后, 自动调整方向保持前行. 结果表明, SNN-CPG 相较于其他两者具有更好的控制精度.

此外, 四足机器人运动过程中偏转角的变化可以体现运动策略的动态调整能力, 如图5所示. 记录四足机器人行进最后的偏转角并且取均值, 如表4所示.

表4 四足机器人偏转角

环境	偏转角/(°)		
	SRL	CPG	SRL-CPG
ENV1	0.07	-0.95	-0.01
ENV2	0.69	23.50	0.16
ENV3	0.75	25.49	0.32

在 ENV1 中, 3 种模型的偏转角在 0°左右, 基本没有发生偏转. 在 ENV2 中, CPG 发生了 50°左右的偏转, 运动方向明显发生了改变; SRL 略微发生了正向偏转, 但是又回到了预期方向; SRL-CPG 一直以预定方向前行. 在 ENV3 中, CPG 出了明显的偏转现象. 这种偏转随着地形复杂性的增加, 偏转的幅度也随之增大; SRL 模型在运动方向上也经历了明显的偏转, 导致其偏离了预定的运动轨迹; SRL-CPG 虽然在初始阶段也出现了轻微的正向偏转, 但经过一段时间的自我调整, SRL-CPG 模型展现出了较强的动态调整能力, 最终回到正确的运动轨迹上.

最后, 为了比较四足机器人行走的稳定性, 可以用机器人运动时的俯仰角变化程度衡量运动稳定性, 其俯仰角变化如图6所示.

将四足机器人运动过程中每 3 个时间步取均值, 利用均值求俯仰角方差, 根据方差大小衡量四足机器人运动过程中行进的稳定性, 方差越大, 则稳定性越低. 俯仰角方差如表5所示.

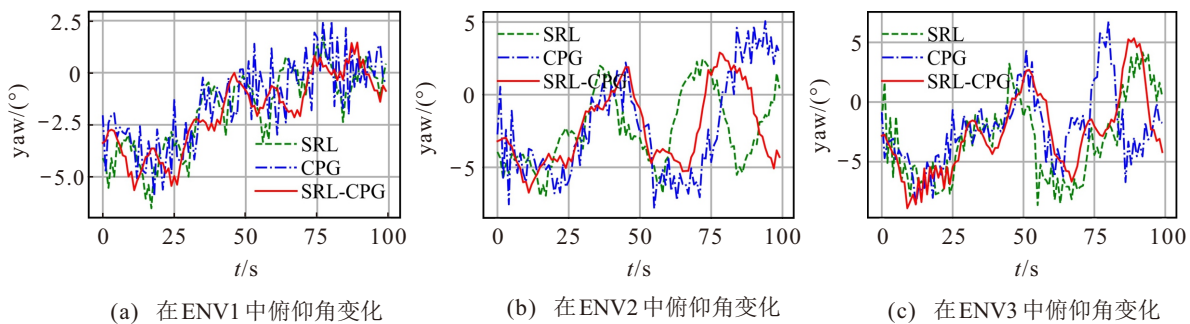


图6 3种地形下四足机器人的俯仰角变化

表5 四足机器人俯仰角方差

环境	模型		
	SRL	CPG	SRL-CPG
ENV1	0.30	1.30	0.16
ENV2	0.15	4.73	0.14
ENV3	0.98	4.58	0.32

在环境复杂程度较低的情况下, CPG 的俯仰角方差与其他两种模型差距不大, 但是环境变复杂之后, CPG 的俯仰角方差急剧变大, 说明 CPG 模型稳定性随着环境变复杂, 稳定性急剧降低. SRL 俯仰角方差在环境最复杂的情况下才会明显上升, 说明其模型稳定性较好, 但随着环境复杂程度升高其稳定性也会明显降低. SRL-CPG 在 3 种环境下, 方差都是最低, 并且其方差并未显著增加, 说明其在运动过程中, 具有最好的运动稳定性.

复杂环境下直行实验表明: SRL-CPG 相较于 SRL 和 CPG, 在简单运动任务下, 具有更好的控制精度, 以及在复杂环境下具有更好的动态调整能力. 此外, 其运动稳定性随环境的影响较小, 能够在复杂环

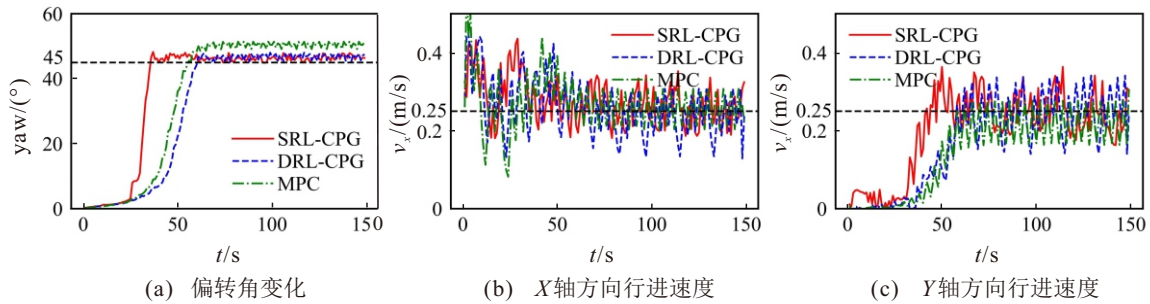


图7 速度控制任务实验结果

在速度稳定之后 ( $t > 100$ ), 计算 3 种方案的平均偏转角、两个方向平均速度, 分别计算相应的百分误差, 作为衡量控制算法控制精度的指标, 计算结果如表 6 所示.

表6 控制偏差对比

误差种类	模型		
	SRL-CPG	DRL-CPG	MPC
偏转角误差/%	1.22	1.89	9.71
$v_x$ 误差/%	1.34	6.48	0.21
$v_y$ 误差/%	5.48	1.32	12.72

图 7 中机器人达到目标方向之后, 就沿着目标方向直行. 图 7 (a) 表示机器人的偏转角变化, 机器人的期望偏转角为  $45^\circ$ , 实验中 SRL-CPG 最先达到偏转角附近, SNN-CPG 的平均偏转角也最接近目标值. 图 7 (b) 和 7(c) 分别表示机器人在 X 和 Y 方向的期望速度, 可以看到 SRL-CPG 能最快到达预期速度.

境下保持稳定的运动状态.

### 3.3 速度控制任务

为了验证 SRL-CPG 模型对更加复杂任务的处理能力, 本文在速度控制任务下测试模型性能. 定义奖励函数

$$\begin{cases} r_t = \xi_{x,y} f(v_{x,y}) + \xi_{\omega_z} f(\omega_z) - \xi_j s_{\text{joint}}, \\ f(x) = \exp\left(-\frac{\|x^{\text{targ}} - x\|^2}{0.25}\right). \end{cases} \quad (13)$$

其中:  $x^{\text{targ}}$  表示期望控制指令, 例如  $v_x^{\text{targ}}$ 、 $v_y^{\text{targ}}$ 、 $\omega_z^{\text{targ}}$  表示在 X 轴和 Y 轴的期望速度方向和偏转角;  $v_x$ 、 $v_y$ 、 $\omega_z$  表示在 X 轴和 Y 轴的实际速度方向;  $s_{\text{joint}}$  表示机器人关节出现异常的惩罚项. 实验过程中  $\xi_{x,y} = 1.0$ ,  $\xi_{\omega_z} = 0.5$ ,  $\xi_j = 0.001$ .

实验中以  $v_x = 0.25$  m/s,  $v_y = 0.25$  m/s, 以及  $\text{yaw} = 45^\circ$  作为四足机器人速度控制目标, 对控制算法进行比较. 为了在速度控制任务下比较 SRL-CPG 的性能, 本文将模型架构与 SRL-CPG 相同的深度强化学习分层控制模型 (DRL-CPG) 以及经典模型预测算法 MPC 作为对照组, 实验结果如图 7 所示.

如表 6 所示, SRL-CPG 和 DRL-CPG 在偏转角误差方面均显著优于 MPC 方法. 此外, SRL-CPG 在速度误差上最优, 表明其在速度控制任务中能够实现更高的控制精度.

实验表明, SRL-CPG 模型能够在保证任务控制精度的情况下使得四足机器人达成速度控制等较为复杂的控制任务, 相较于其他控制方法具有更高的控制精度, 并且 SRL-CPG 在完成速度控制任务时, 具有更快的调整速度.

### 3.4 能耗比较

脉冲神经网络模拟了生物神经元的工作原理, 仅在接收到输入信号时才激活, 这种事件驱动的计算方式使得 SNN 在无信号输入时保持低功耗状态. 与传统的人工神经网络相比, 后者在没有输入时也保持活跃状态. 此外, 在运算过程中后者只有加法运算而没有乘法运算, 因此 SNN 在进行运算时更加节能. SRL-CPG 策略的主要优势之一是 SNN 在运行

过程中的低能耗特性. SRL-CPG 和 DRL-CPG 均根植于神经网络进行计算, 具有相似的理论基础, 可以进行能耗对比计算. 然而, MPC 的能耗计算机制与神经网络存在显著差异, 本文仅对 SRL-CPG 和 DRL-CPG 的能耗进行比较分析. SRL-CPG 策略中能量消耗的计算比较复杂, 在信号经过编码器、解码器、CPG 三个模块中具有浮点型的加法和乘法操作, 而在 SNN 模块中只进行浮点型的加法操作. 对于 DRL-CPG, 其每一个模块都要进行浮点型乘法和加法的操作. 模型能耗计算公式如下 (在 45 nm 的芯片组技术中,  $E_{MAC} = 4.6 \text{ pJ}$ ,  $E_{AC} = 0.9 \text{ pJ}$ ):

$$E_{\text{model}} = E_{\text{MAC}} \cdot FL_{\text{Other}} + E_{\text{AC}} \cdot FL_{\text{SNN}}. \quad (14)$$

直接将需要加法的模块分别与乘法模块和加法模块相加, 乘以对应的耗能, 即可得到结果. 按照脉冲发放周期的步长分别进行对比, 结果如表 7 所示.

表7 SRL-CPG 和 DRL-CPG 的能耗比较

周期步长	能耗/nJ		能量节省占比/%
	DRL-CPG	SRL-CPG	
$T = 1$	31.42	4.84	84.60
$T = 2$	31.42	9.41	60.01
$T = 3$	31.42	19.76	37.12

如表 7 所示, 在种群编码过程中, 脉冲发放周期步长  $T$  变化, SRL-CPG 的模型节能占比也会随之变化, 但是在 3 种时间步长下, SRL-CPG 模型的能耗都要低于 DRL-CPG. 在本文选取的模型中  $T = 1$ , 利用 SRL-CPG 模型能降低 84.6% 的控制能耗.

## 4 结论

本文基于脉冲强化学习和 CPG 提出了一种分层运动控制算法来改善四足机器人运动控制中的控制性能和能耗问题. 所提出的方法以 LIF 神经元建立脉冲神经网络, 并以混合训练的架构构造 SRL 模型; 通过引入 CPG 模型, 构建具有步态生成功能的低级控制器, 同时以 SRL 接收状态反馈, 调整 CPG 的输入参数, 保证了控制策略的控制性能. 仿真实验结果表明, 所提出算法能在保持控制性能的前提下具有更快的动态调整速度, 并且在能耗上相较于传统的深度学习方法拥有巨大的优势.

本文所做的工作是将 SRL 和 CPG 相结合应用于四足机器人运动控制的初步探索, 仍然有很多不足. 例如, 为了便于控制, 本文简化了四足机器人模型, 未考虑其偏转髋关节和腰部关节对运动的影响; 本文在设置控制任务时主要考虑的是四足机器人的行走运动, 未能在跳跃、翻转这种更加考验四足机器人的平衡性能的运动任务中进行测试. 此外, 本文提

出的分层控制架构并未在四足机器人实物模型进行测试. 这些都将是下一步的研究方向.

## 参考文献 (References)

- [1] Kumar. Neural networks a classroom approach[M]. New York: Mcgraw Hill Higher Education, 2002.
- [2] Marder E, Bucher D. Central pattern generators and the control of rhythmic movements[J]. Current Biology, 2001, 11(23): R986-R996.
- [3] 董豪, 杨静, 李少波, 等. 基于深度强化学习的机器人运动控制研究进展[J]. 控制与决策, 2022, 37(2): 278-292.  
(Dong H, Yang J, Li S B, et al. Research progress of robot motion control based on deep reinforcement learning[J]. Control and Decision, 2022, 37(2): 278-292.)
- [4] 孙辉辉, 胡春鹤, 张军国. 移动机器人运动规划中的深度强化学习方法[J]. 控制与决策, 2021, 36(6): 1281-1292.  
(Sun H H, Hu C H, Zhang J G. Deep reinforcement learning for motion planning of mobile robots[J]. Control and Decision, 2021, 36(6): 1281-1292.)
- [5] Yao Q F, Wang J L, Wang D L, et al. Hierarchical terrain-aware control for quadrupedal locomotion by combining deep reinforcement learning and optimal control[C]. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. Prague, 2021: 4546-4551.
- [6] Kulkarni S, Simon S P, Sundareswaran K. A spiking neural network(SNN) forecast engine for short-term electrical load forecasting[J]. Applied Soft Computing, 2013, 13(8): 3628-3635.
- [7] Glackin C, McDaid L, Maguire L, et al. Implementing fuzzy reasoning on a spiking neural network[C]. Lecture Notes in Computer Science. Berlin, 2008: 258-267.
- [8] Nichols E, McDaid L J, Siddique N. Biologically inspired SNN for robot control[J]. IEEE Transactions on Cybernetics. 2013, 43(1): 115-128.
- [9] Batllori R, Laramee C B, Land W, et al. Evolving spiking neural networks for robot control[J]. Procedia Computer Science, 2011, 6: 329-334.
- [10] Tang G Z, Kumar N, Yoo R, et al. Deep reinforcement learning with population-coded spiking neural network for continuous control[C]. Conference on Robot Learning. PMLR, 2021: 2016-2029.
- [11] Zhang D Z, Zhang T L, Jia S C, et al. Multi-scale dynamic coding improved spiking actor network for reinforcement learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(1): 59-67.
- [12] Wu X D, Ma S G. CPG-based control of serpentine locomotion of a snake-like robot. Mechatronics[J]. 2010,

- 20(2): 326-334.
- [13] Liu C J, Chen Y F, Zhang J Q, et al. CPG driven locomotion control of quadruped robot[C]. 2009 IEEE International Conference on Systems, Man and Cybernetics. San Antonio, 2009: 2368-2373.
- [14] rder E, Bucher D. Central pattern generators and the control of rhythmic movements[J]. *Current Biology*, 2001, 11(23): 986-996.
- [15] Kimura H, Fukuoka Y, Katabuti H. Mechanical design of a quadruped “Tekken3&4” and navigation system using laser range sensor[C]. International Symposium on Robotics, DOI: 10.1103/PhysRevB.72.024502.
- [16] Kuramoto Y. Self-entrainment of a population of coupled non-linear oscillators[C]. International Symposium on Mathematical Problems in Theoretical Physics. Kyoto, 1975: 420-422.
- [17] Hu Y H, Liang J H, Wang T M. Parameter synthesis of coupled nonlinear oscillators for CPG-based robotic locomotion[J]. *IEEE Transactions on Industrial Electronics*. 2014, 61(11): 6183-6191.
- [18] Liu C J, Xia L, Zhang C Z, et al. Multi-layered CPG for adaptive walking of quadruped robots[J]. *Journal of Bionic Engineering*. 2018, 15(2): 341-355.
- [19] Auddy S, Magg S, Wermter S. Hierarchical control for bipedal locomotion using central pattern generators and neural networks[C]. The 9th International Conference on Development and Learning and Epigenetic Robotics. Oslo, 2019: 13-18.
- [20] Bellegarda G, Ijspeert A. CPG-RL: Learning central pattern generators for quadruped locomotion[J]. *IEEE Robotics and Automation Letters*. 2022, 7(4): 12547-12554.
- [21] Lele A S, Fang Y, Ting J, et al. Learning to walk: Spike based reinforcement learning for hexapod robot central pattern generation[C]. The 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems. Genova, 2020: 208-212.

### 作者简介

肖云发 (1999-), 男, 硕士, 主要研究方向为脉冲强化学习, E-mail: [xiao\\_yunfa@mail.dhu.edu.cn](mailto:xiao_yunfa@mail.dhu.edu.cn);

韩芳 (1981-), 女, 教授, 博士, 主要研究方向为神经动力学和智能控制, E-mail: [yadiahah@dhu.edu.cn](mailto:yadiahah@dhu.edu.cn);

王青云 (1972-), 男, 教授, 博士, 主要研究方向为神经和智能动力学与控制, E-mail: [qingyunwang@buaa.edu.cn](mailto:qingyunwang@buaa.edu.cn).