



浙江大学
ZHEJIANG UNIVERSITY

基于动作生成的足式机器人运动控制

浙江大学博士研究生学位申请与毕业设计论文答辩

答辩人：邵烨程

指导老师：杨卫，王宏涛

ANYmal



Spot



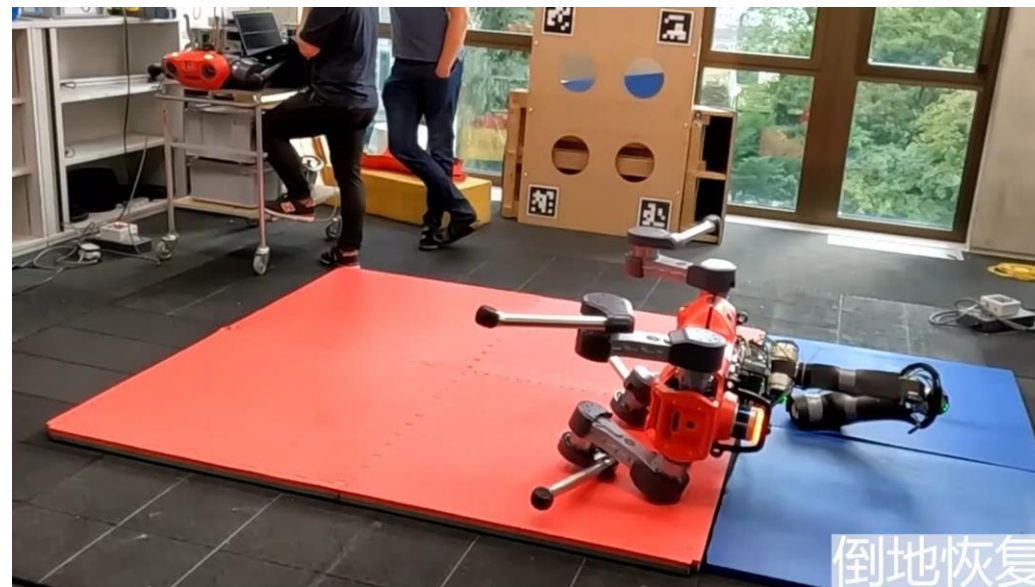
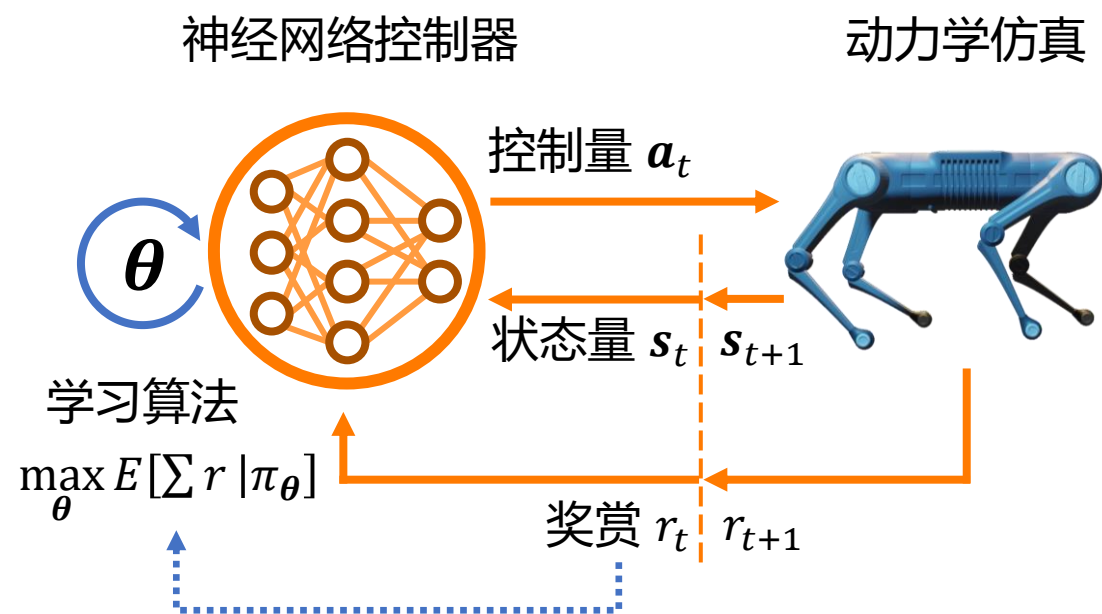
宇树 B1



绝影 Lite3

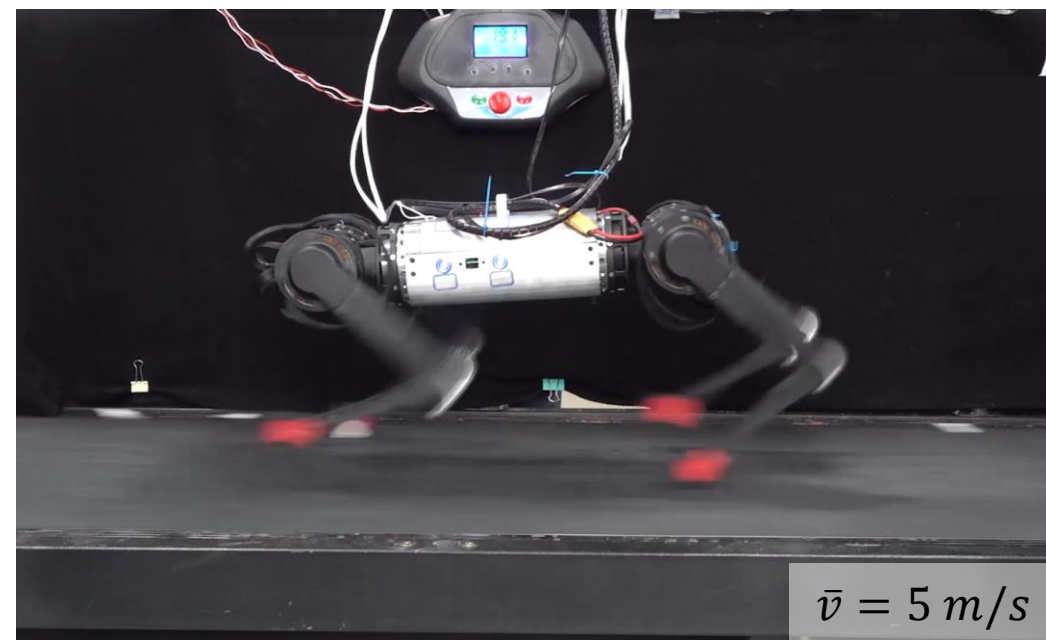
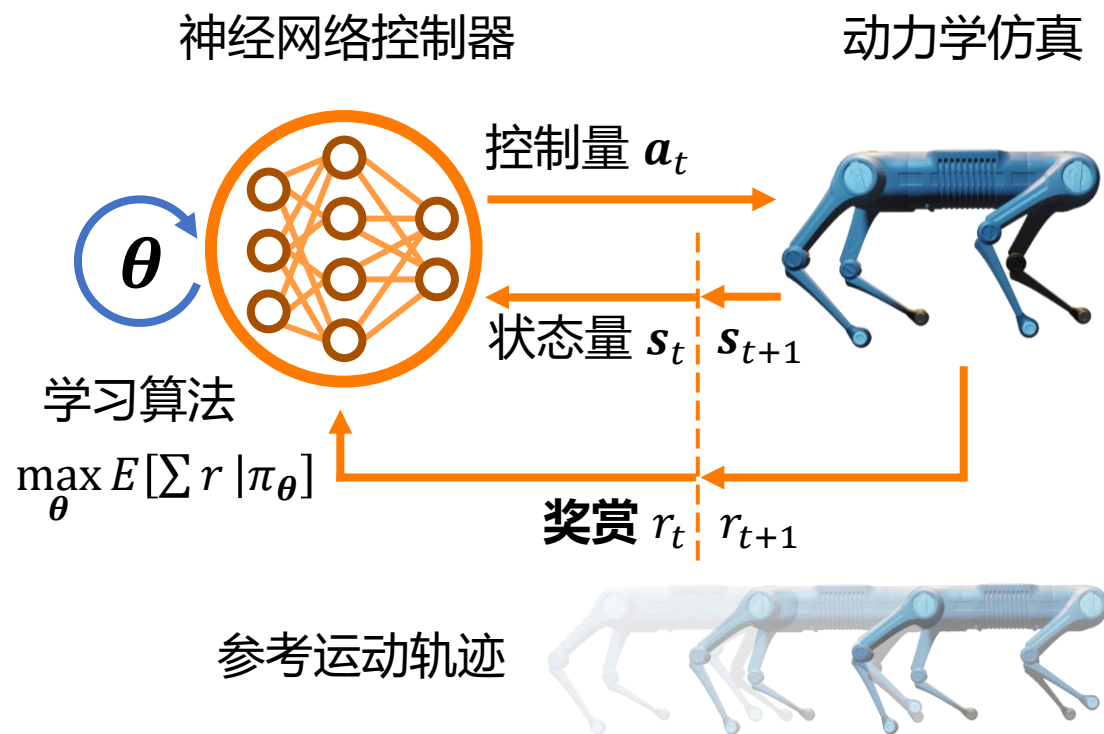


基于强化学习的方法大量应用于传统控制器难以处理的**复杂问题**



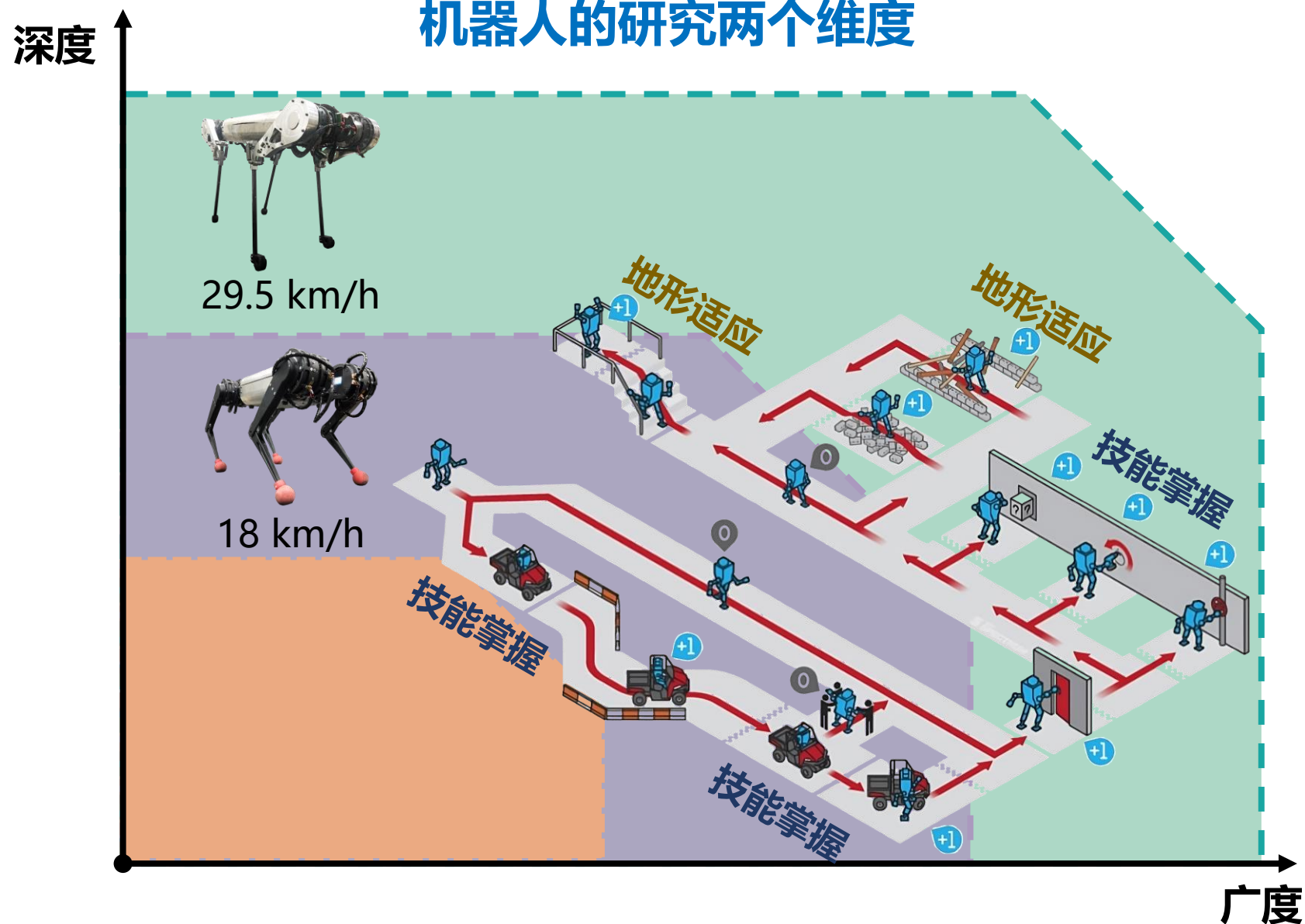
- 使用**神经网络**将机器人的状态量映射至控制量，结构简单，计算**效率高**
- 学习算法以**最大化奖励**为目标，基于动力学**仿真**数据**优化**网络参数，并迁移至**实物**

奖励函数是强化学习中最重要的一部分，对控制器有很大影响



- 借助**参考运动**轨迹，使用**动作模仿**简化**奖励函数**设计
- 实现了小尺寸机器人 **5 m/s** 的高速奔跑，**超越**基于模型的控制方法

机器人的研究两个维度



□ 机器人**深度**方向的研究提高机器人的**作业能力上限**

□ 机器人**广度**方向的研究推进了机器人的**落地与应用**

□ 掌握多种技能

□ 适应不同环境

多任务学习



- 模型同时学习**多个**有限任务的能力
- 任务冲突、遗忘等问题

泛化迁移

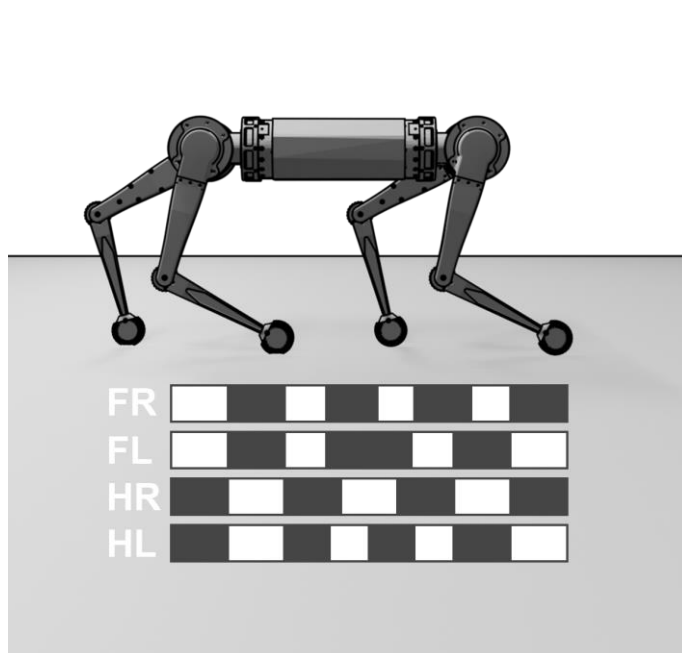


- 模型在训练集**以外**的样本中的表现
- 训练环境与测试环境不匹配等问题

多任务学习与**泛化迁移**是机器学习中的两个重要研究领域

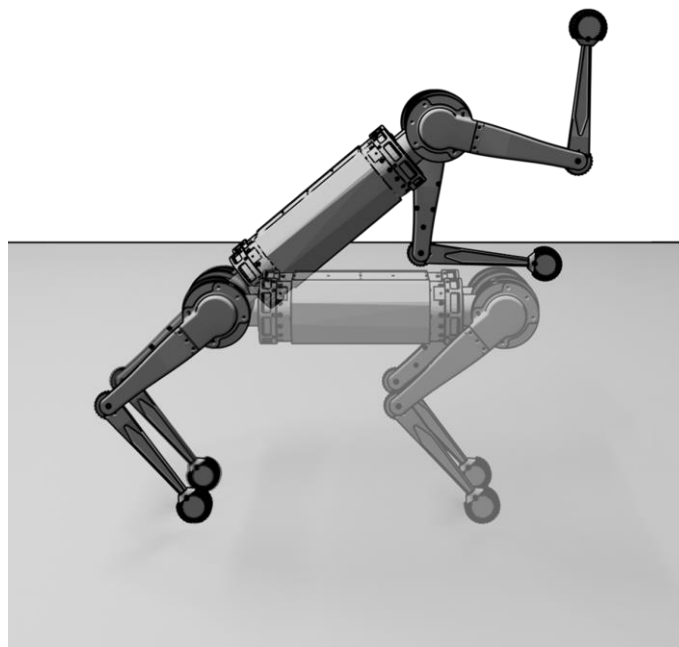
多任务学习

步态切换



- 基于**相位**引导的控制器
- 分析**引导信号**的作用

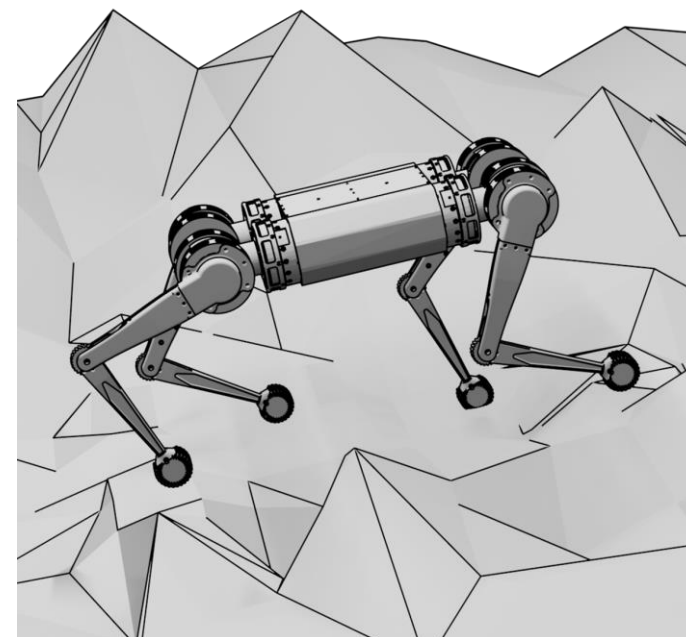
复杂动作



- 使用**运动学**轨迹引导
- 探索参考轨迹**时间关联**

泛化迁移

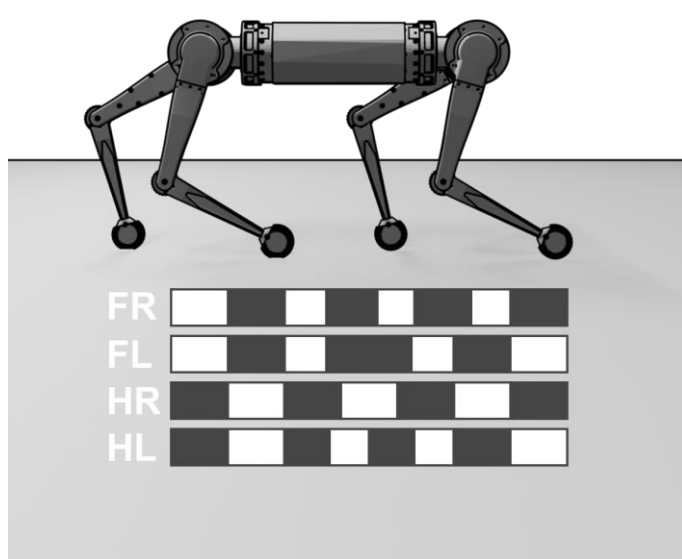
地形适应



- 使用**落脚点**作为引导信号
- 定量**描述地形**差异

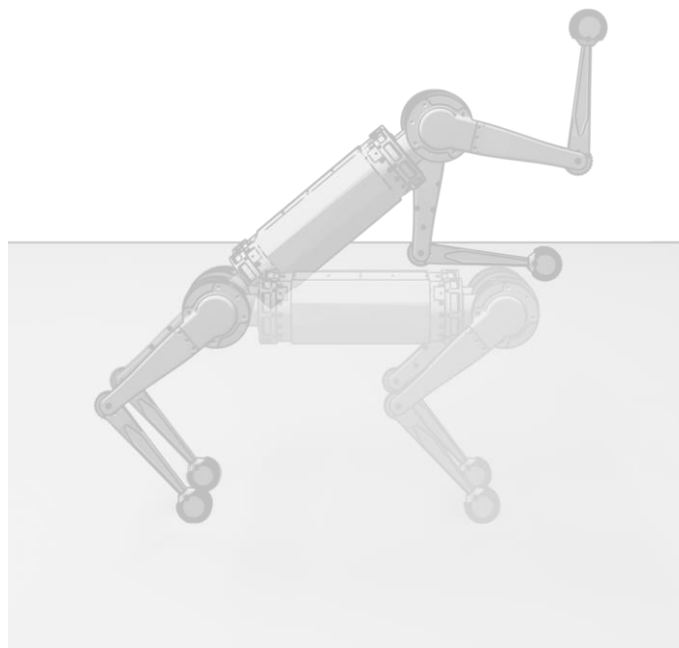
多任务学习

步态切换



- 基于**相位**引导的控制器
- 分析**引导信号**的作用

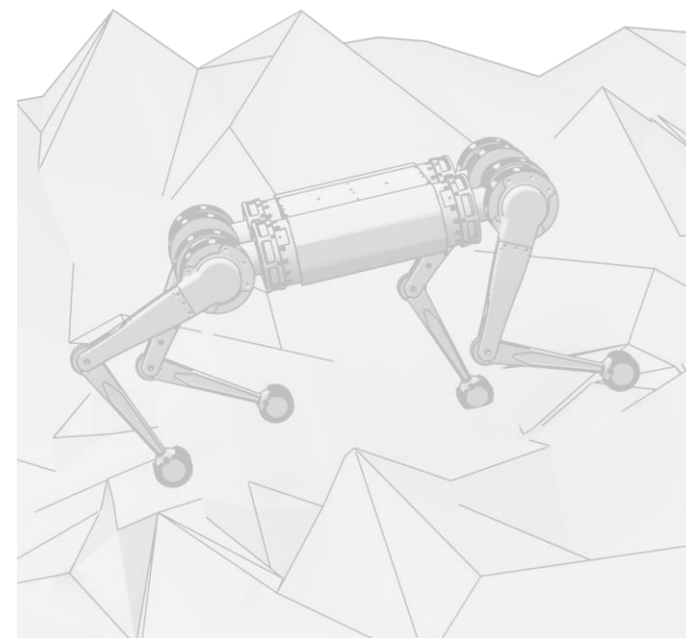
复杂动作



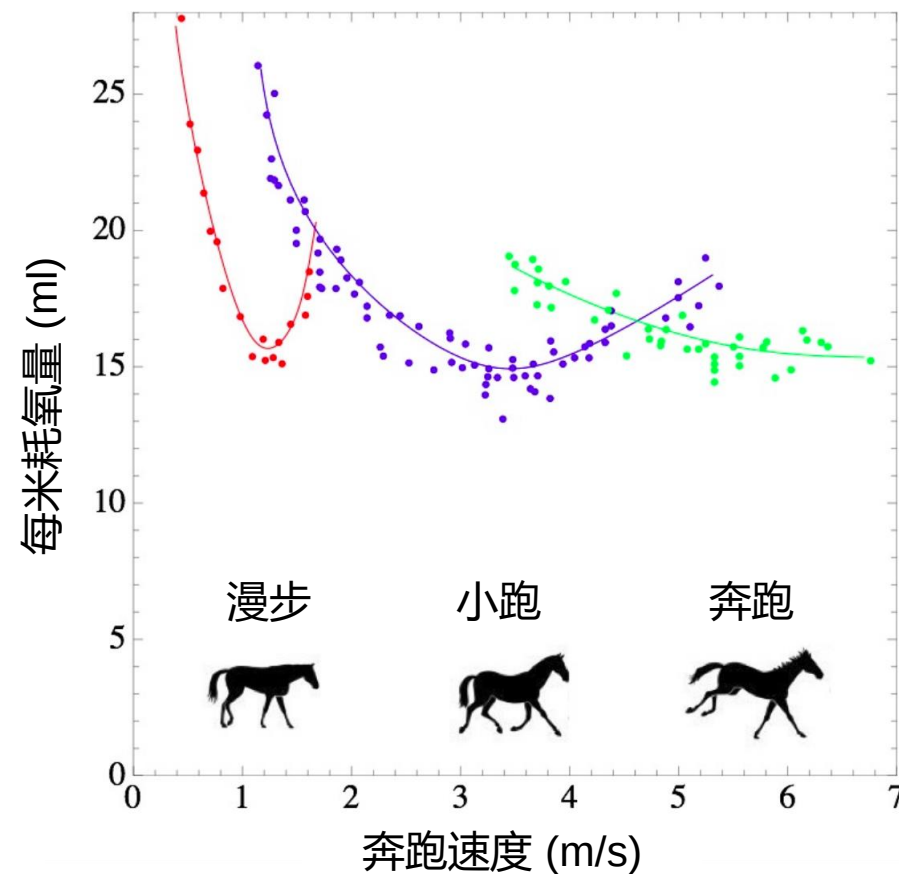
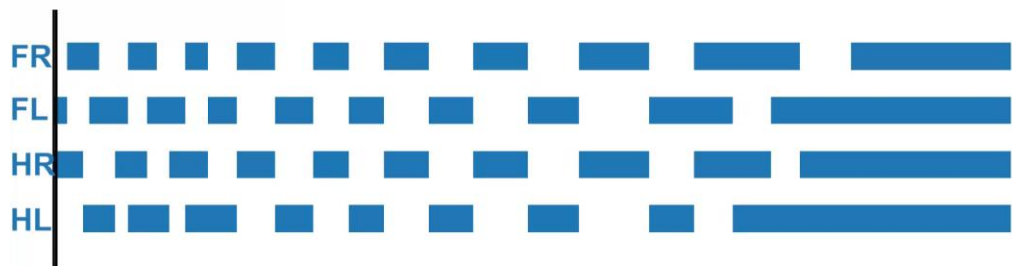
- 使用**运动学**轨迹引导
- 探索参考轨迹**时间关联**

泛化迁移

地形适应

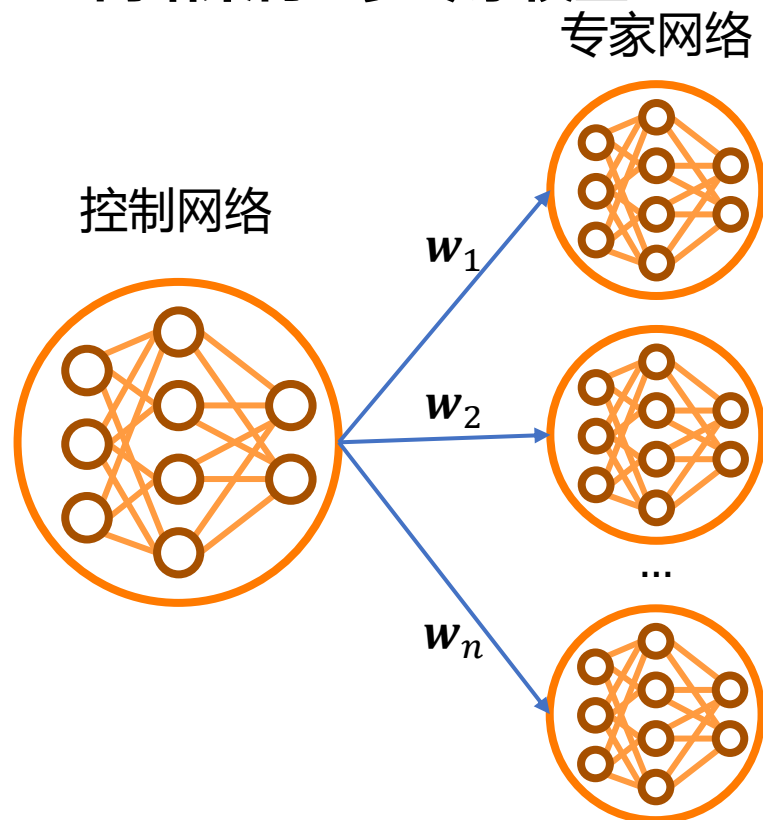


- 使用**落脚点**作为引导信号
- 定量**描述地形**差异



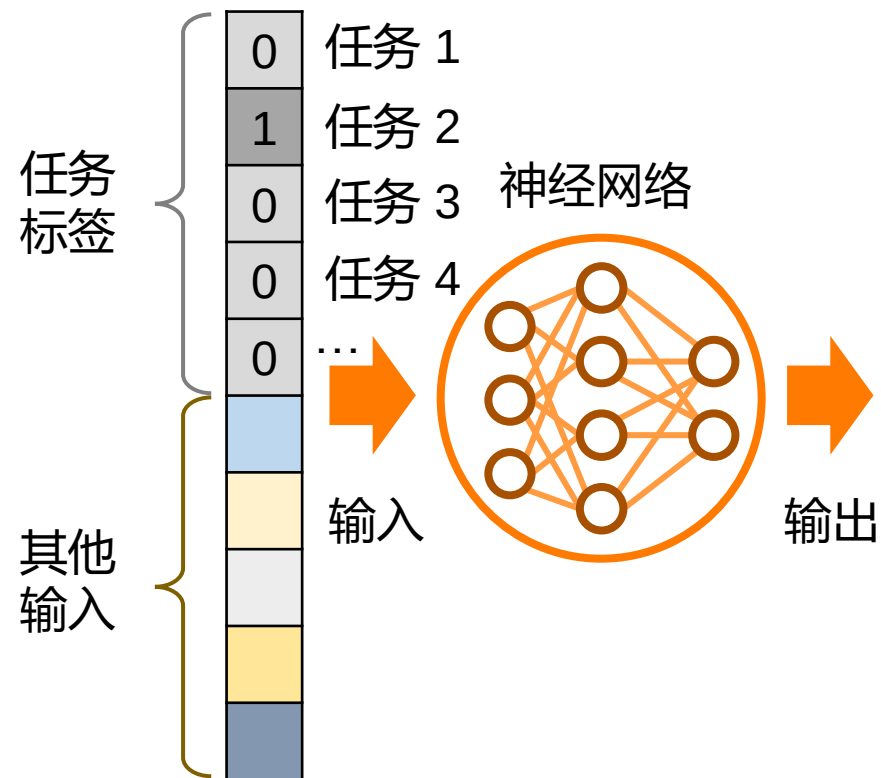
- 动物根据速度、地形、机动性等需求，选择最合适的**步态**并自由地进行**步态切换**
- 同时学习多种步态的运动控制是一个典型的**多任务学习**问题

网络架构：多专家模型



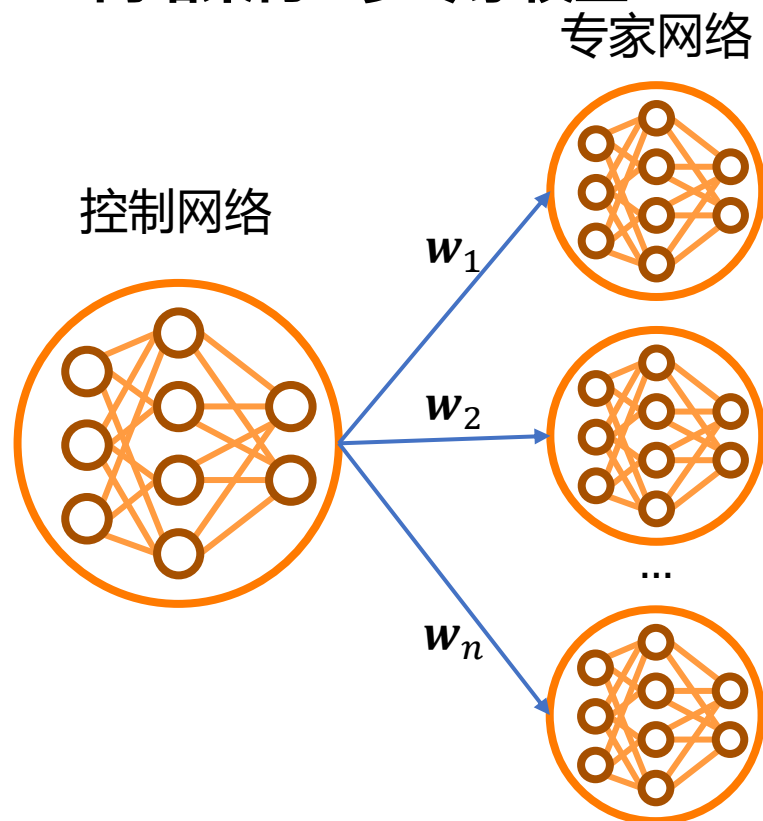
- 通过调控**专家网络**权重实现多任务学习
- 模型参数多、**计算效率低**

输入设计：多任务标签

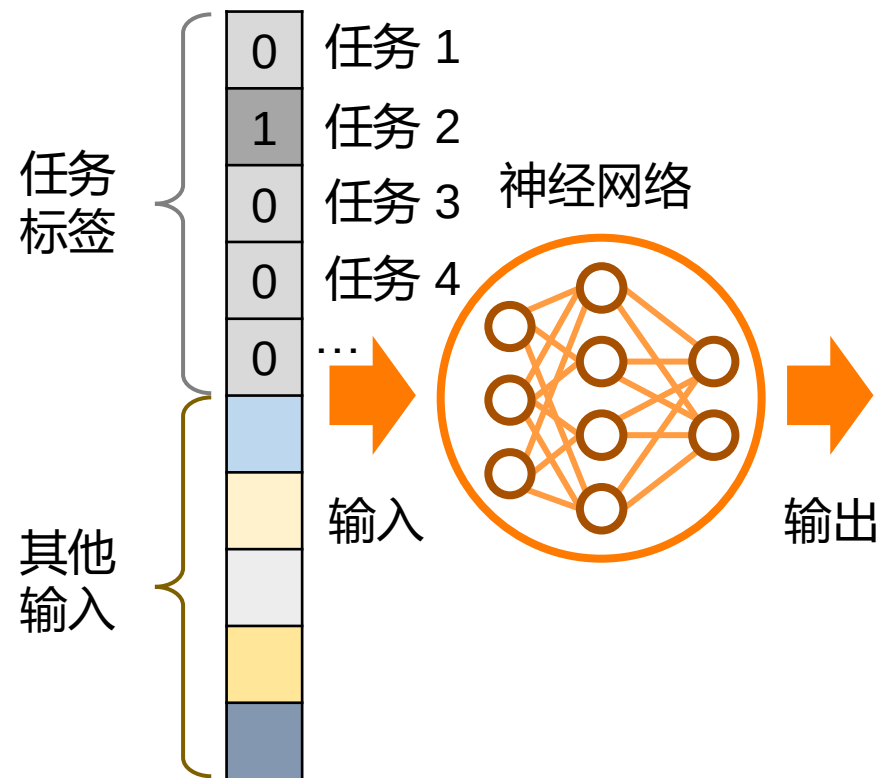


- 通过改变**输入**区分不同任务
- 模型简单，计算效率高
- 标签设计**扩展性差**

网络架构：多专家模型

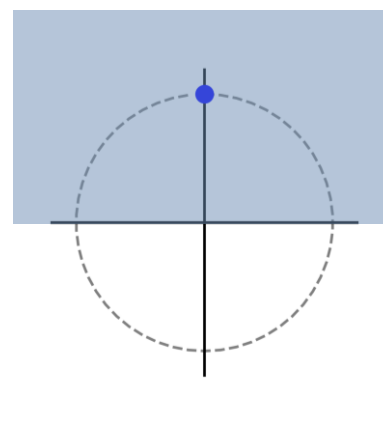
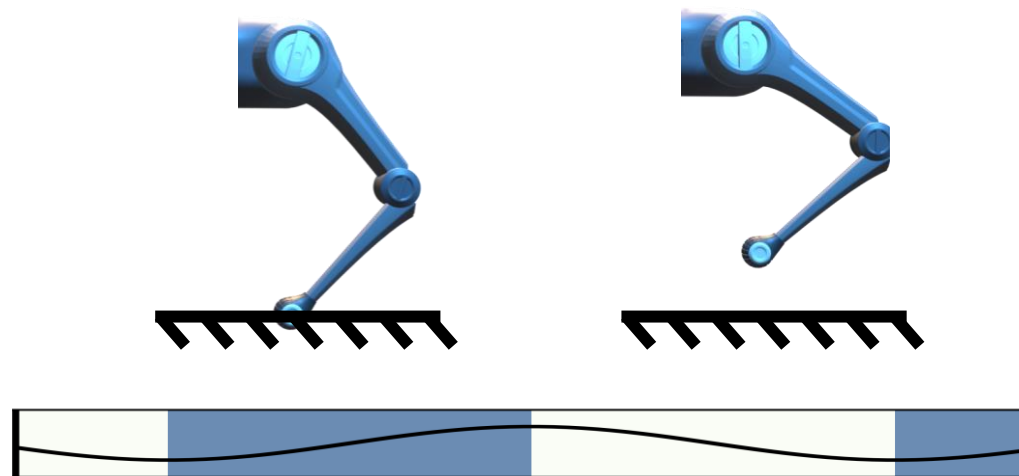
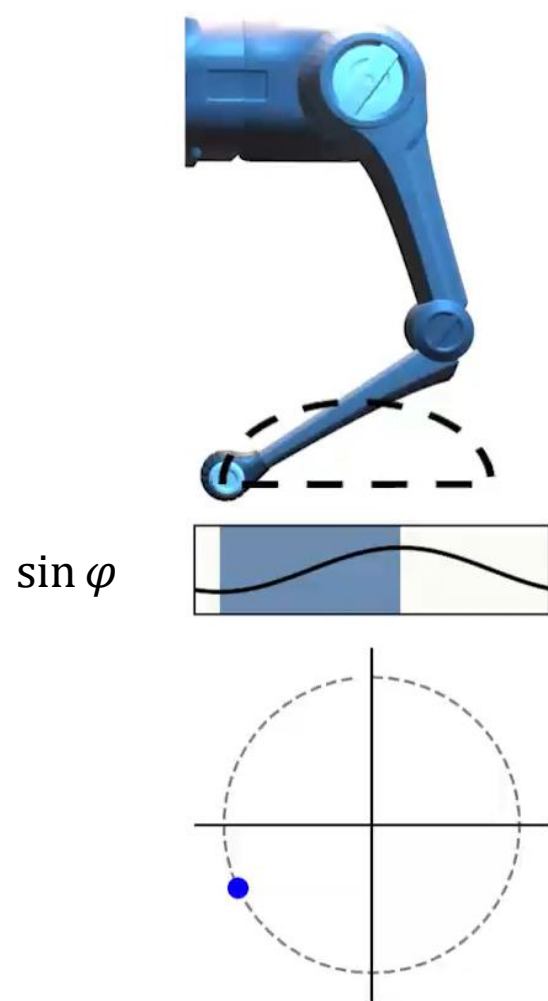


输入设计：多任务标签

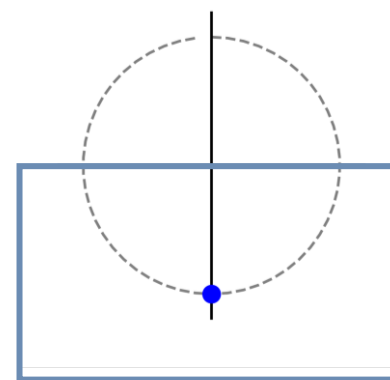


对于足式机器人的运动控制，是否存在一种更加简单、合理的任务表示方法？

使用**相位**描述一条腿的**周期性运动**

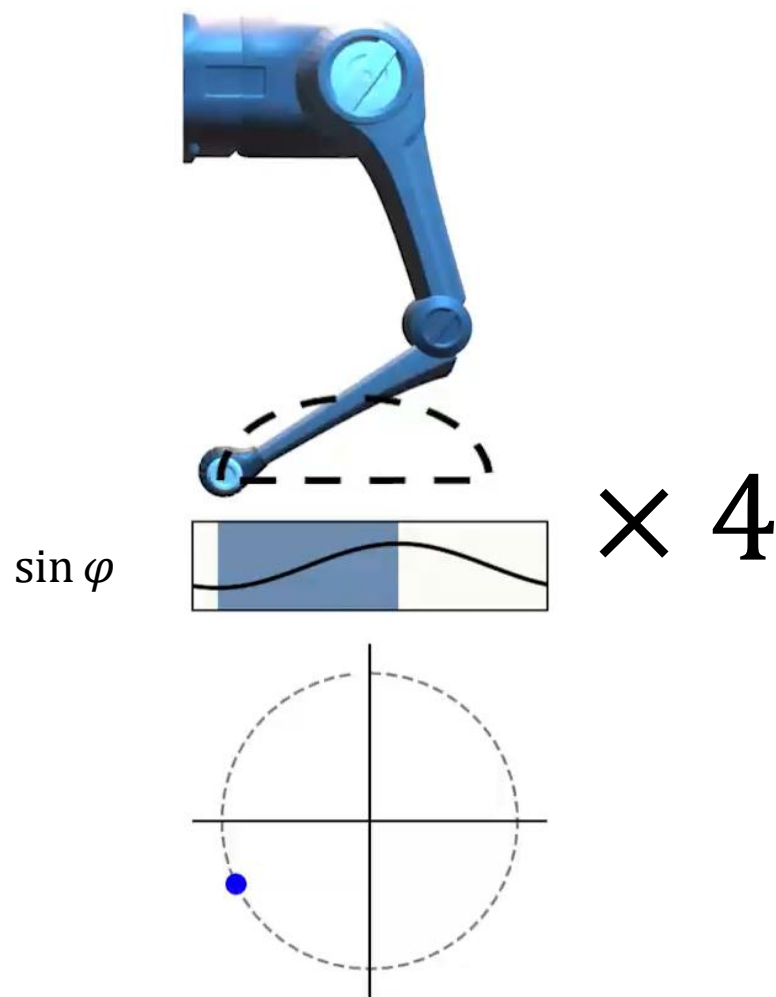


支撑相

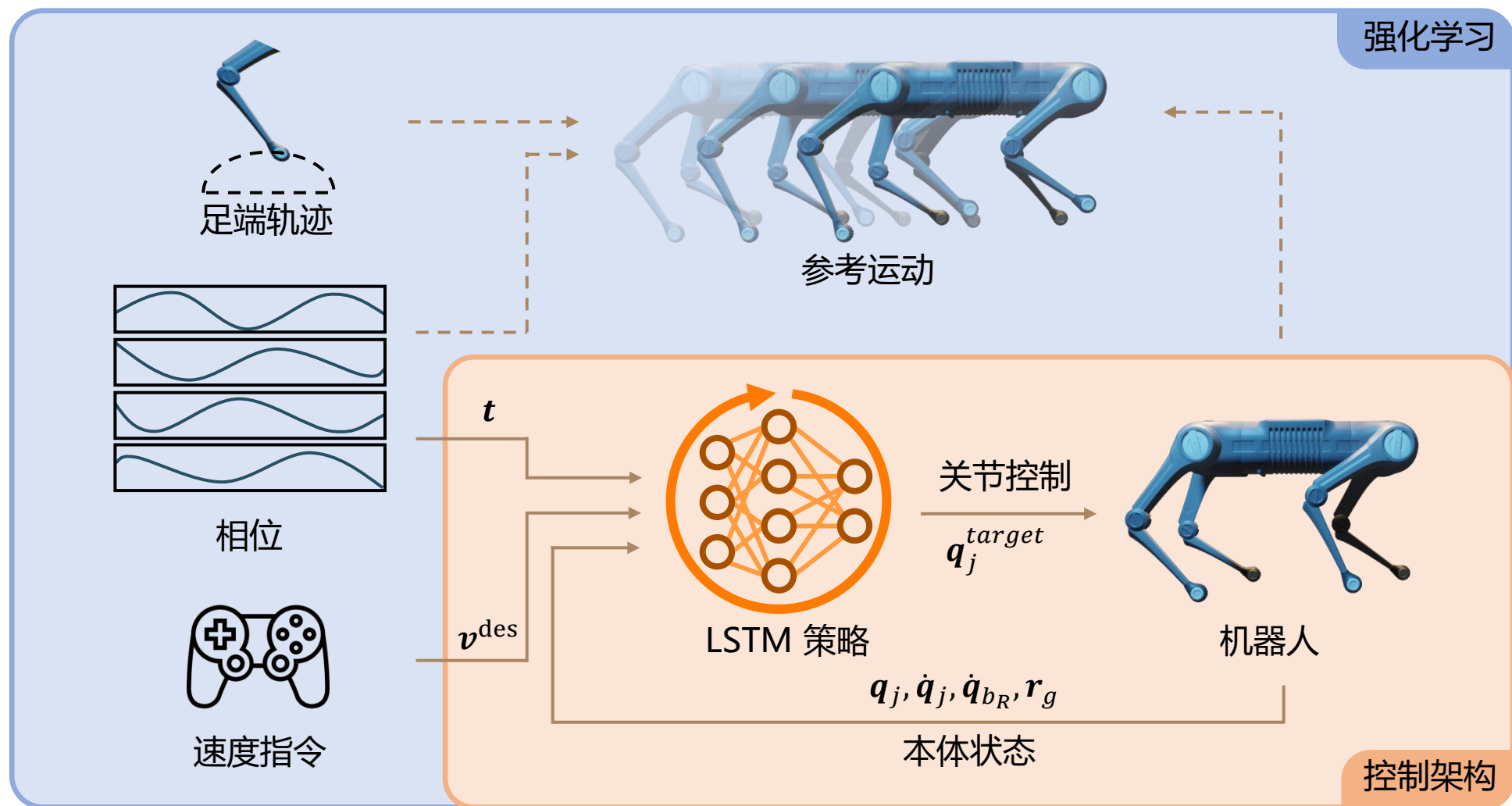


摆动相

使用**多个相位**的组合描述**任意步态**

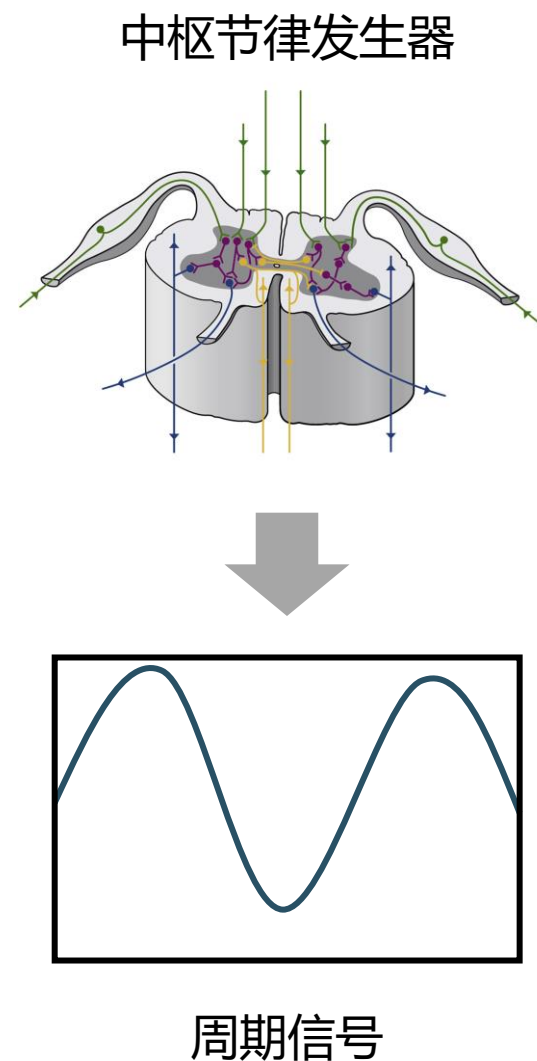
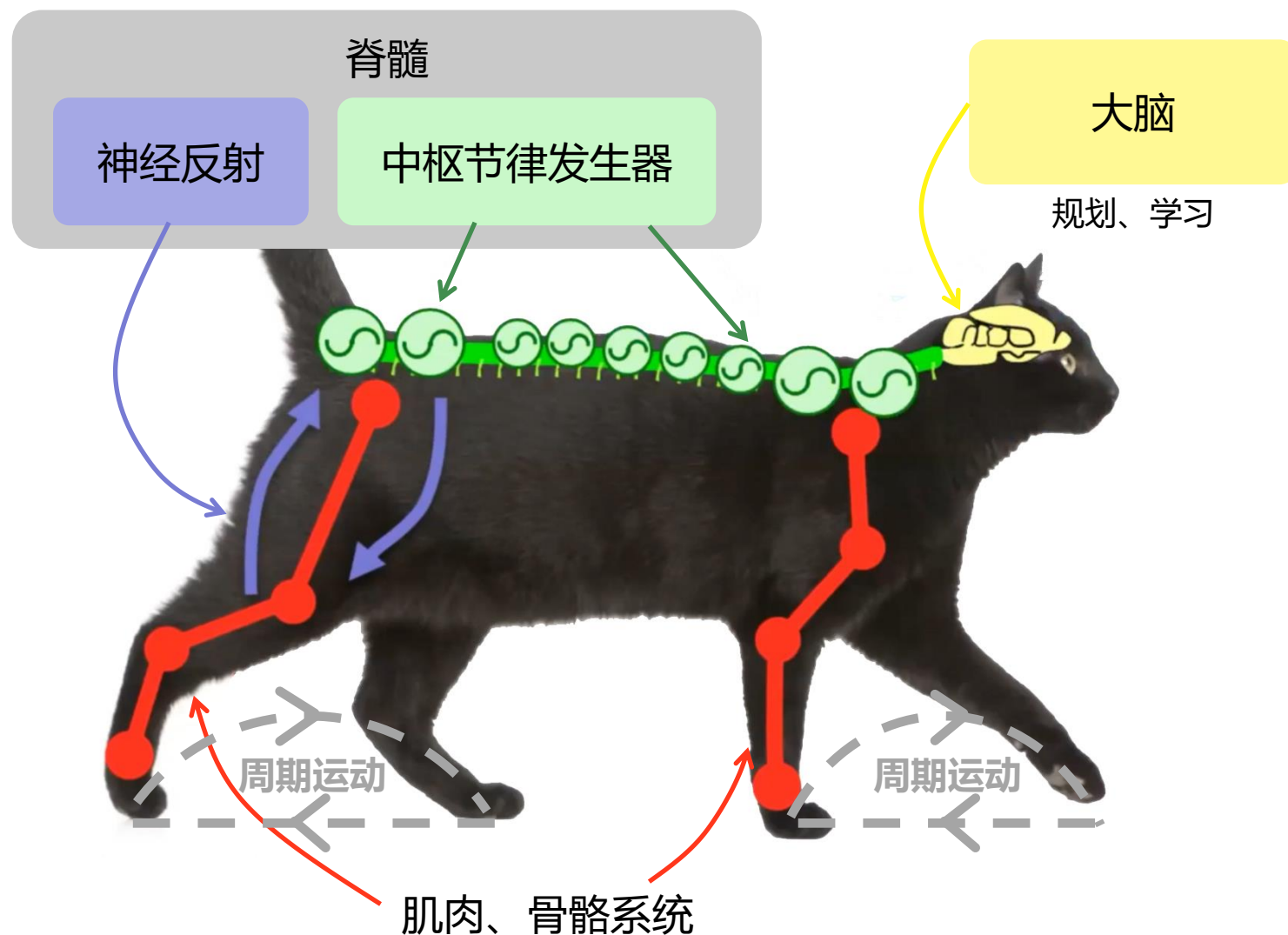


四个相位可以描述**所有**四足运动步态



- 使用周期变化的**相位**作为对于不同步态的**统一**描述
- 将任务从学习**多种步态**的运动简化为学习相位与腿往复摆动之间的**对应关系**

在生物学中，中枢节律发生器 (Centre Pattern Generator, CPG) 被认为是**周期运动**的来源



在机器人研究中，中枢节律发生器 (CPG) 的**数学模型**被用于产生不同步态的运动

单个 Hopf 振子的**平面动力学系统**

$$\begin{cases} \dot{\rho}_0 = \alpha(\mu^2 - \rho^2)\rho_0 - \omega\rho_1 \\ \dot{\rho}_1 = \alpha(\mu^2 - \rho^2)\rho_1 + \omega\rho_0 \\ \gamma = \frac{\pi}{\beta T(e^{-b\rho_1} + 1)} + \frac{\pi}{(1 - \beta)T(e^{b\rho_1} + 1)} \end{cases}$$

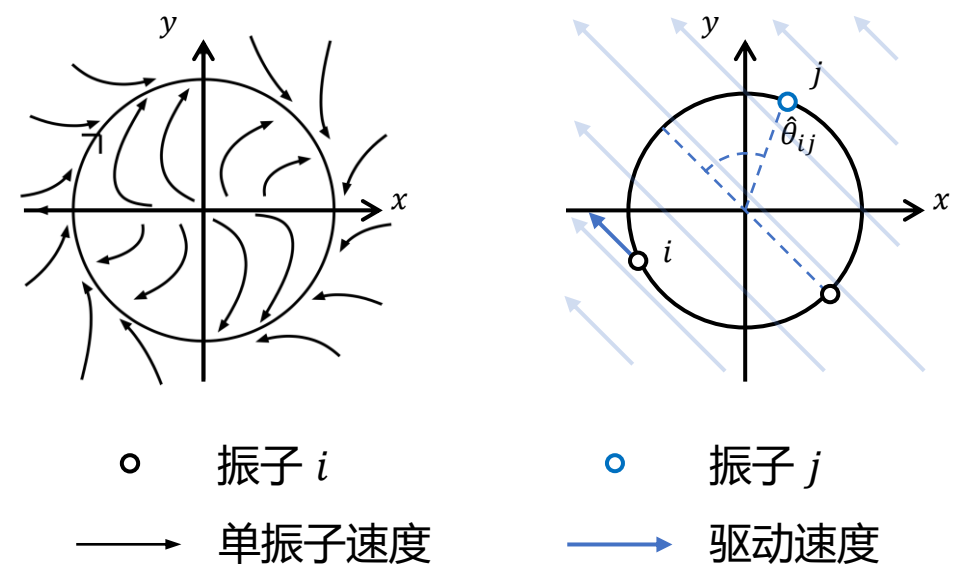
多个 Hopf 振子间的**相互作用**

$$\dot{\rho}^{(i)} = f(\rho^{(i)}) + \delta \sum_{i=1}^4 R_{ij} \rho^{(j)}$$

驱动速度

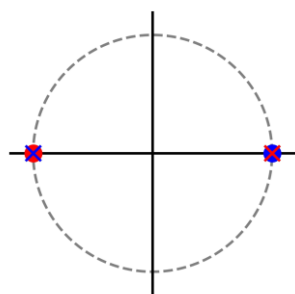
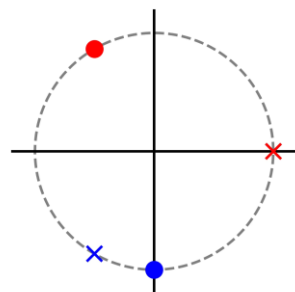
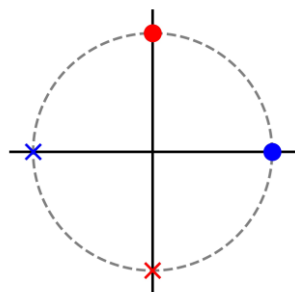
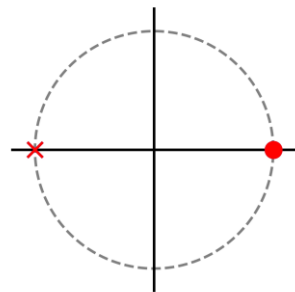
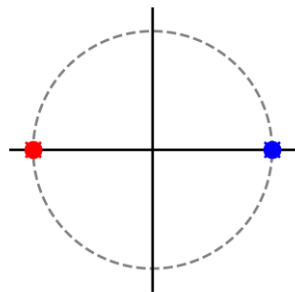
相互作用矩阵

$$R_{ij} = \begin{bmatrix} \cos \theta_{ij} & -\sin \theta_{ij} \\ \sin \theta_{ij} & \cos \theta_{ij} \end{bmatrix}$$
$$\theta_{ij} = \varphi_j - \varphi_i$$



步态	相位差
漫步 (walk)	(0, π, 0.5π, 1.5π)
踱步 (pace)	(0, π, 0, π)
小跑 (trot)	(0, π, π, 0)
奔跑 (bounding)	(0, 0, π, π)

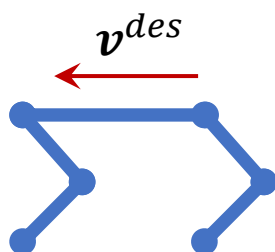
步态切换



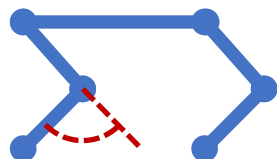
- FL
- FR
- × HL
- × HR

中枢节律发生器可用于生成多种四足与三足的**常见步态**

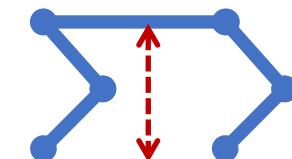
子奖赏函数	函数形式
速度跟随	$r = 0.5\exp(-2.0\ \Delta v\ ^2) + 0.5\exp(-2.0\ \Delta \omega\ ^2)$
关节模仿	$r = 0.25\exp(-2.0\ \Delta p\ ^2) + 0.75\exp(-0.01\ \Delta \dot{p}\ ^2)$
高度保持	$r = \exp(-80\ \Delta h\ ^2)$
姿态保持	$r = \exp(-80\ \Delta r_g\ ^2)$
接触相序	$r = \exp(-\ v_{\text{touch}}\ ^2 + \ \Delta h\ ^2)$



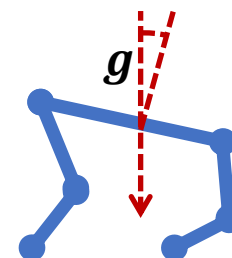
速度跟随



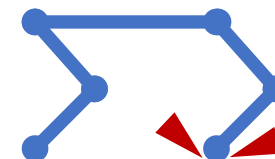
关节模仿



高度保持

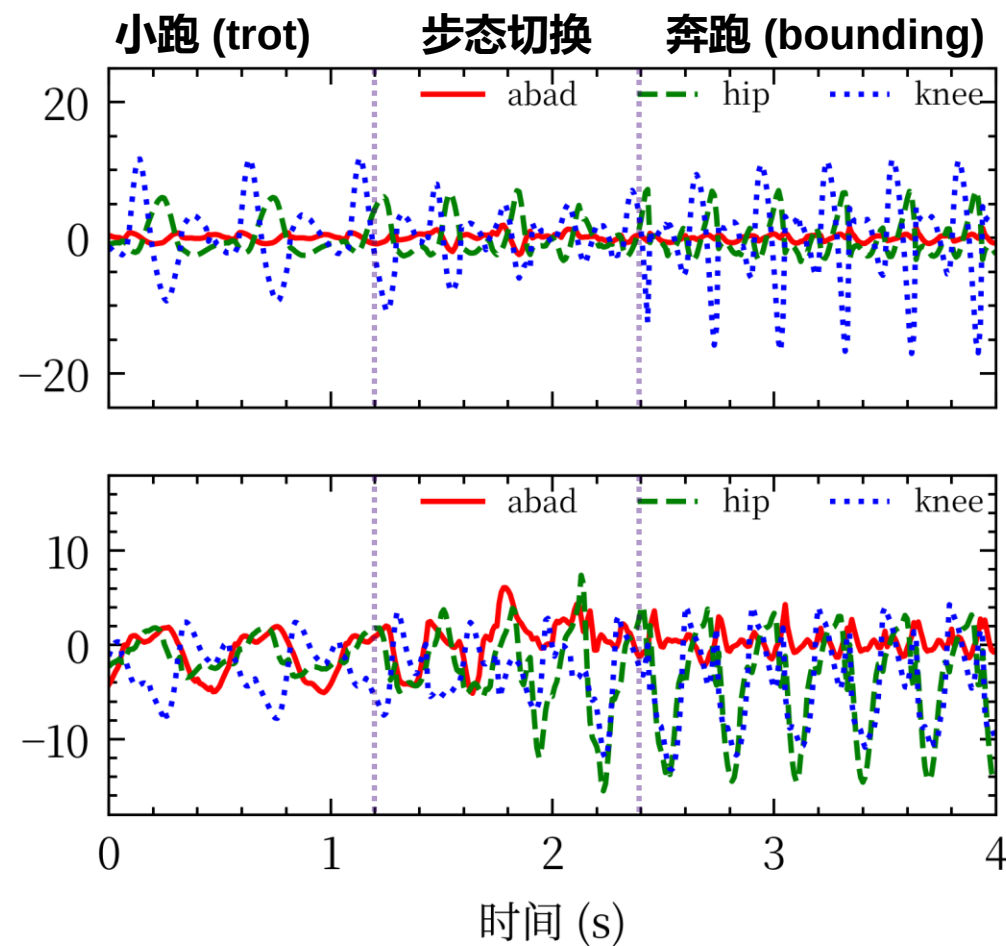
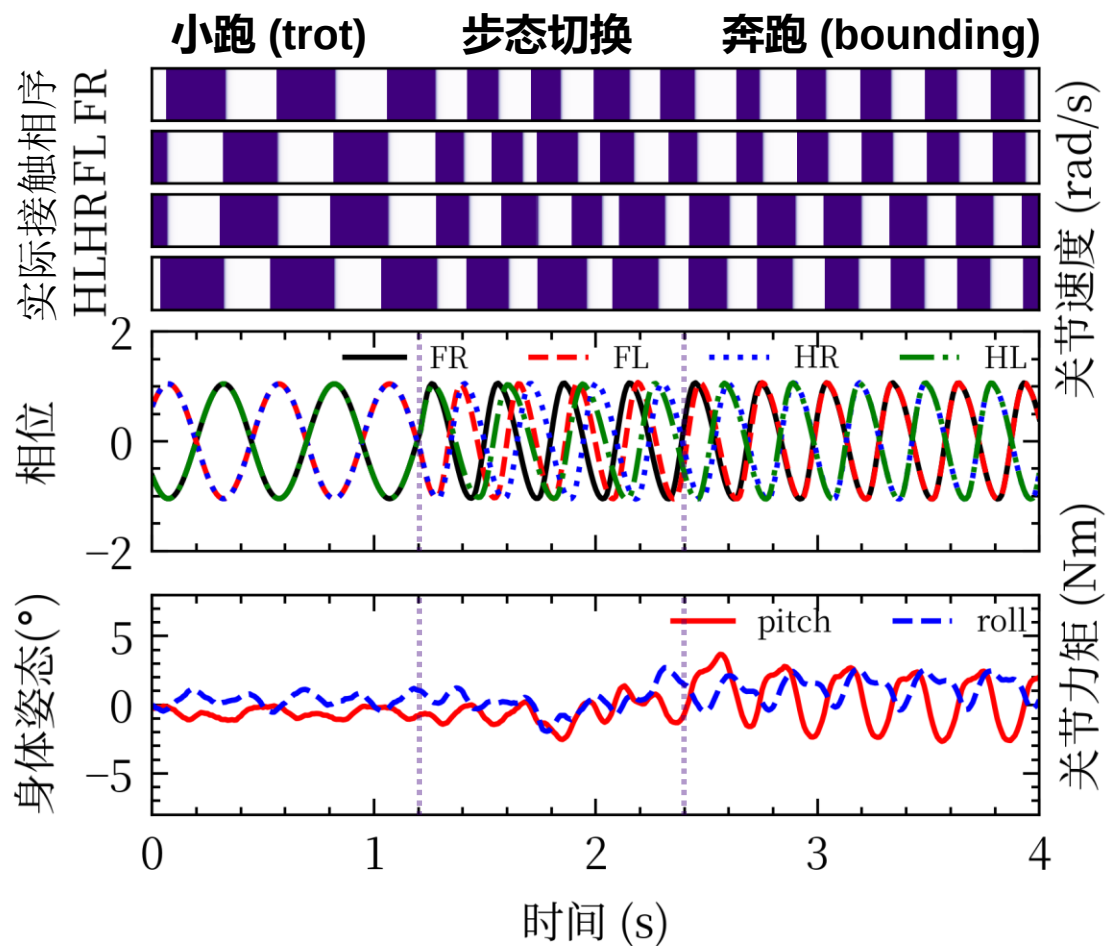


姿态保持



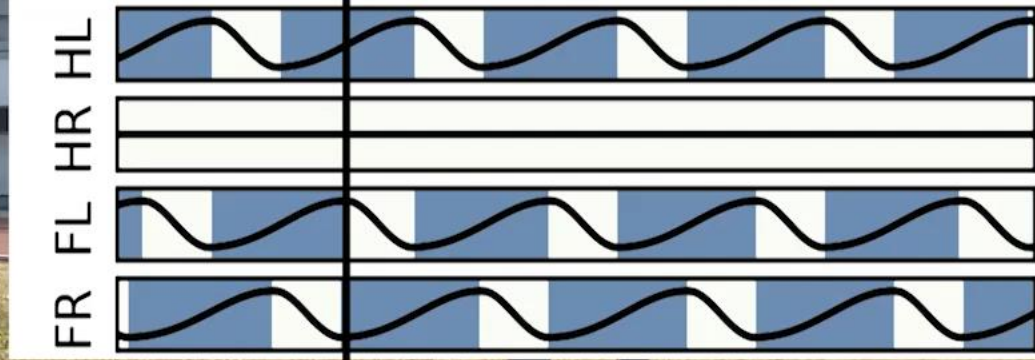
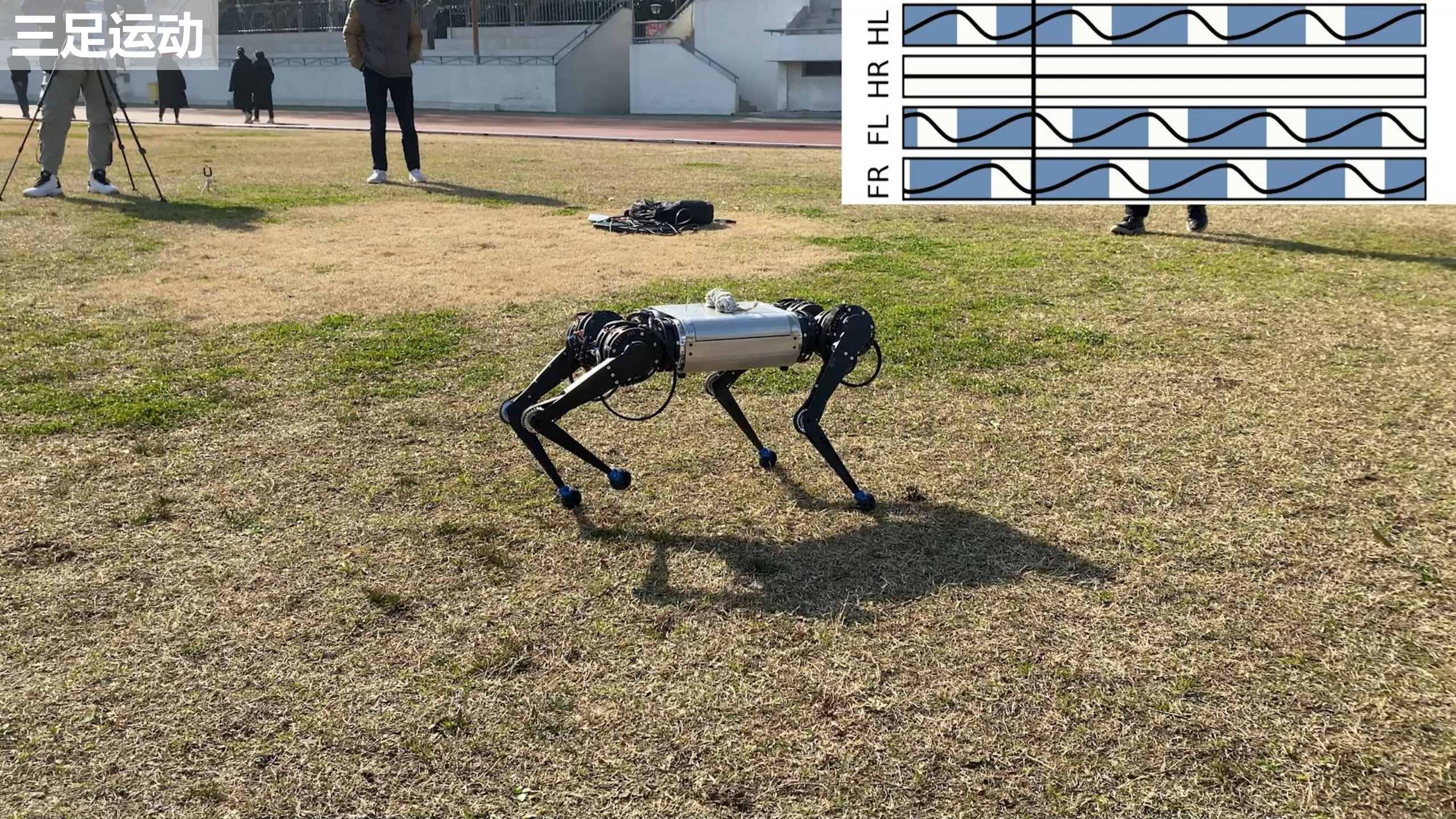
接触相序

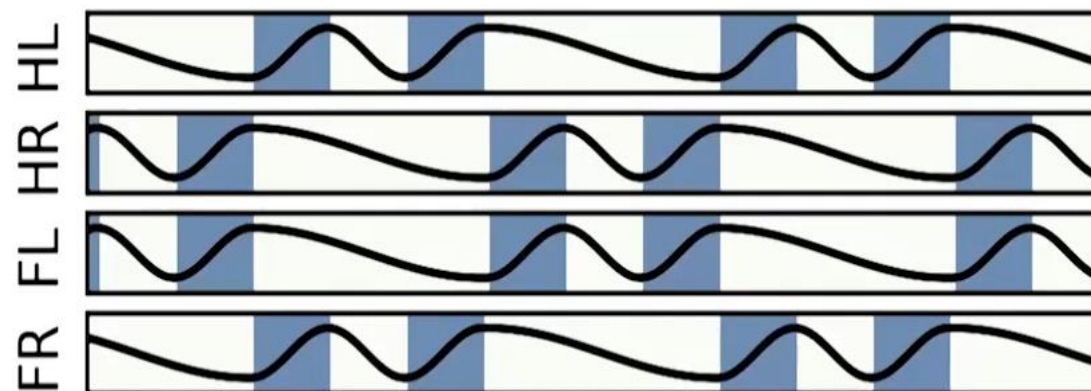
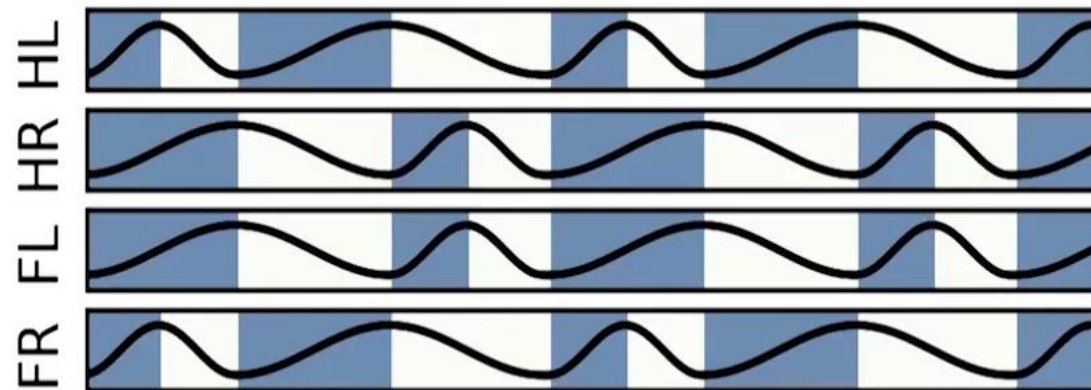
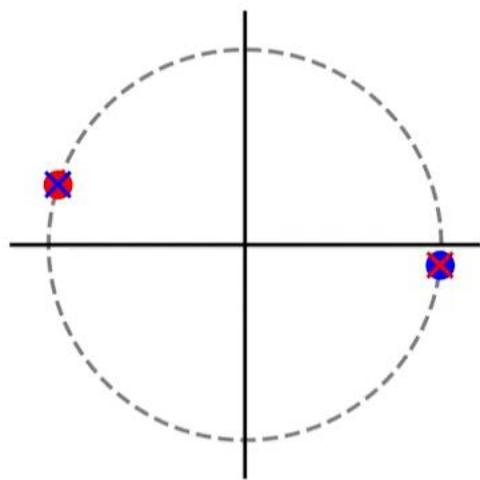
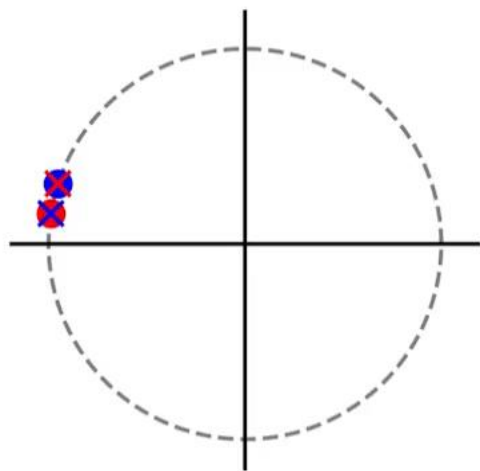
多步态运动与步态切换



- 在步态切换过程中，四只脚的**接触状态**能够基本**遵循**相位**引导**
- 步态切换过程中，机器人身体**姿态平稳**，电机的速度、力矩**变化平滑**

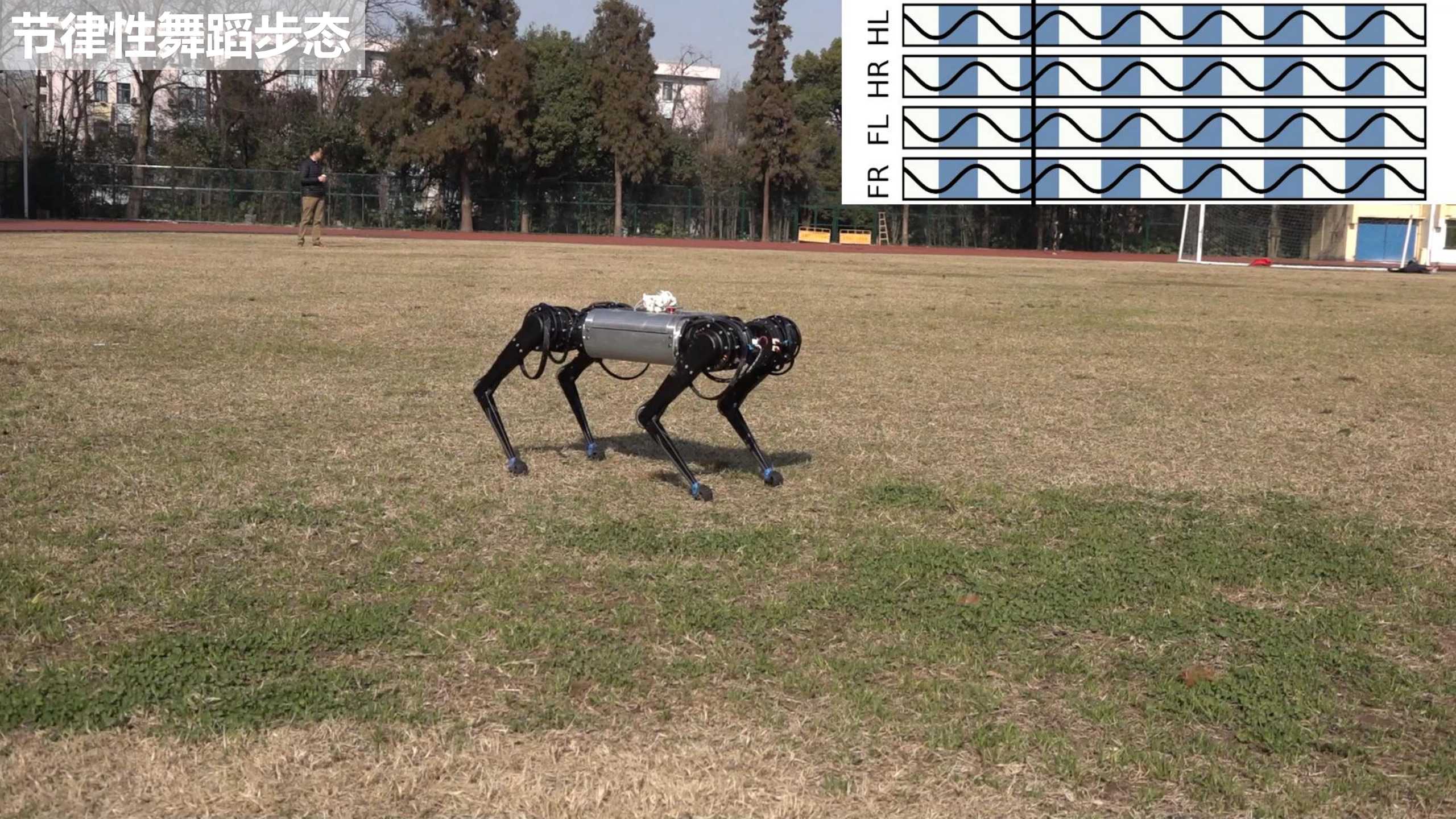
三足运动



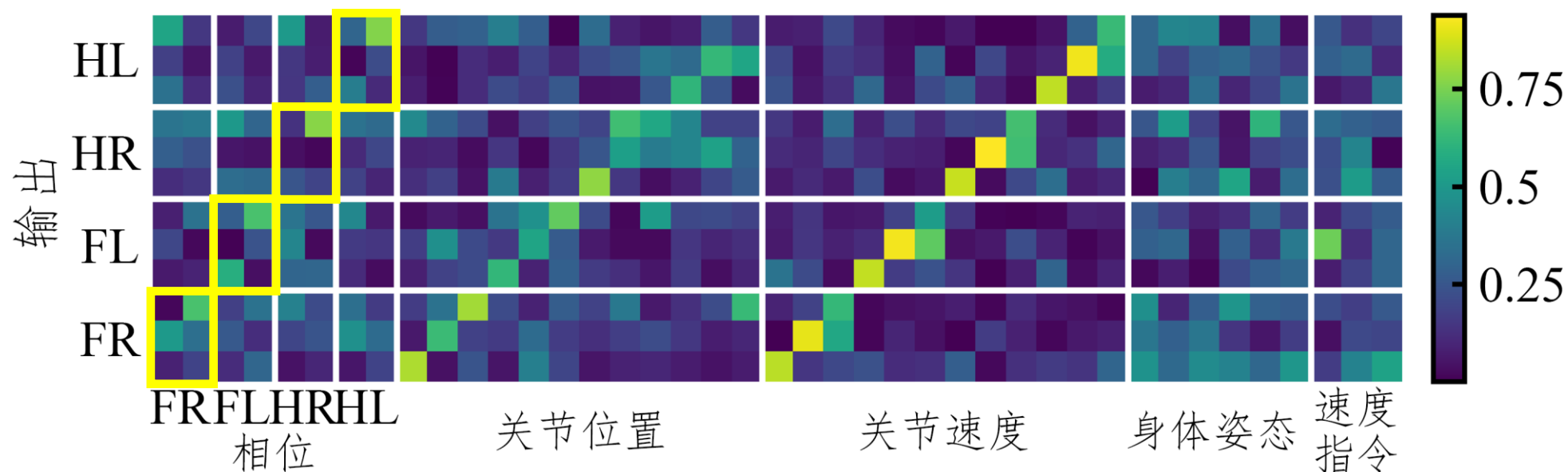


人工设计函数**形式灵活**，可生成等多种不同的**节律性步态**

节律性舞蹈步态



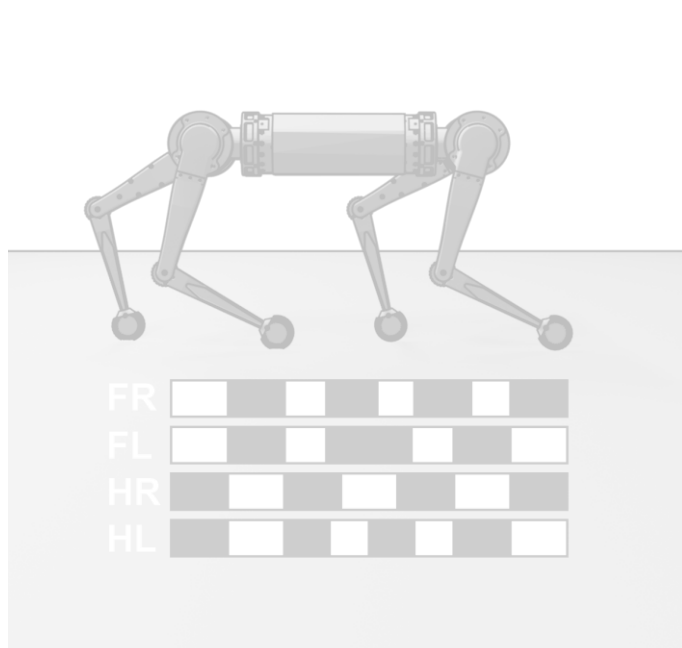
神经网络输出对输入的**敏感程度**: $S_{ij} = E \left[\left| \frac{\partial y_i}{\partial x_j} \right| \right]$



- 输出中每条腿的**控制量**对相应的**相位**的敏感程度高于其他腿
- 控制器学会了根据**相位引导**控制每条腿的运动，从而**实现多步态运动与切换**

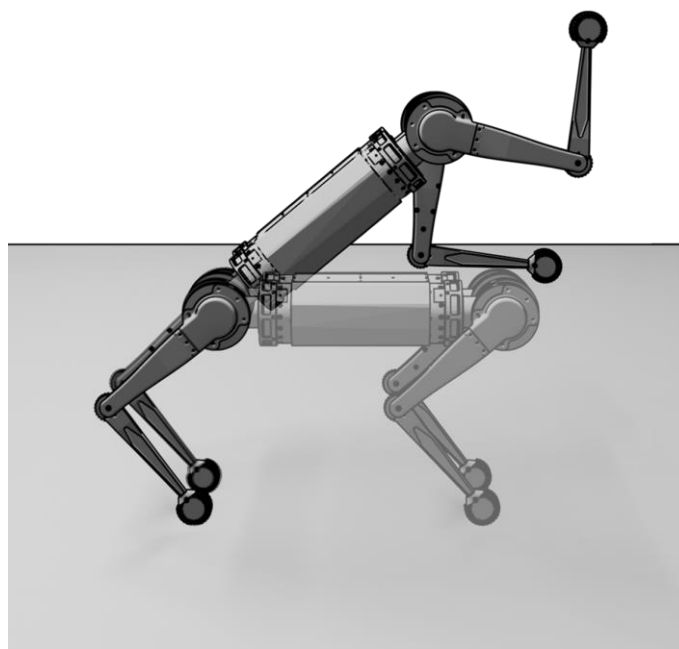
多任务学习

步态切换



- 基于**相位**引导的控制器
- 分析**引导信号**的作用

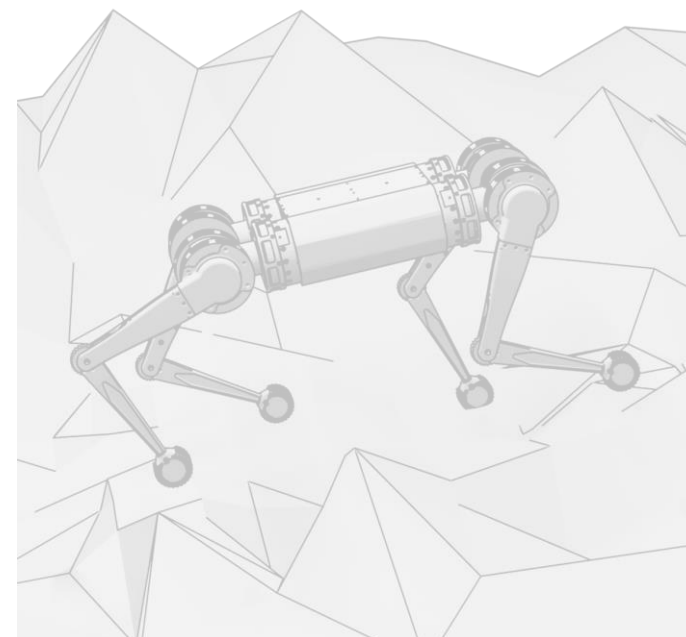
复杂动作



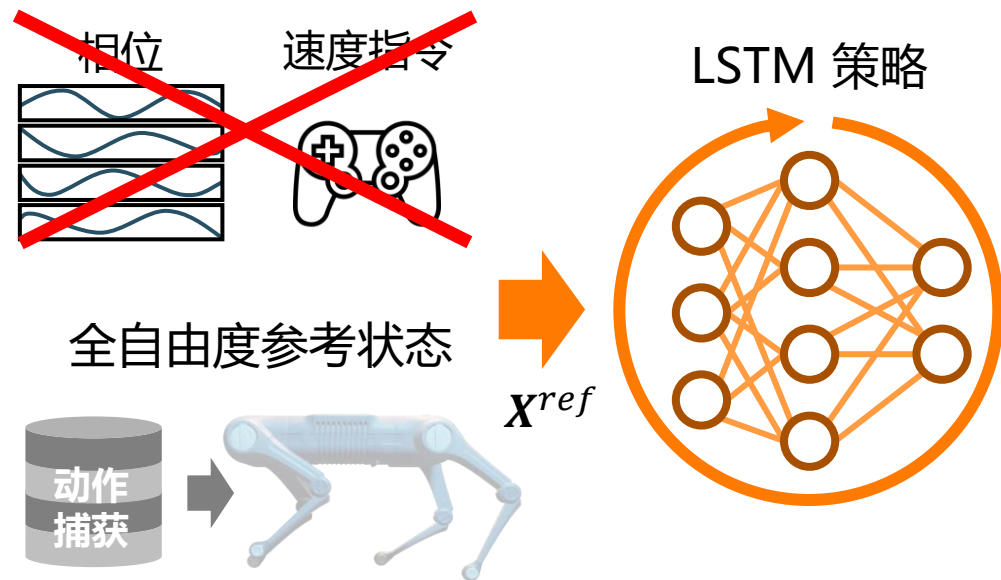
- 使用**运动学**轨迹引导
- 探索参考轨迹**时间关联**

泛化迁移

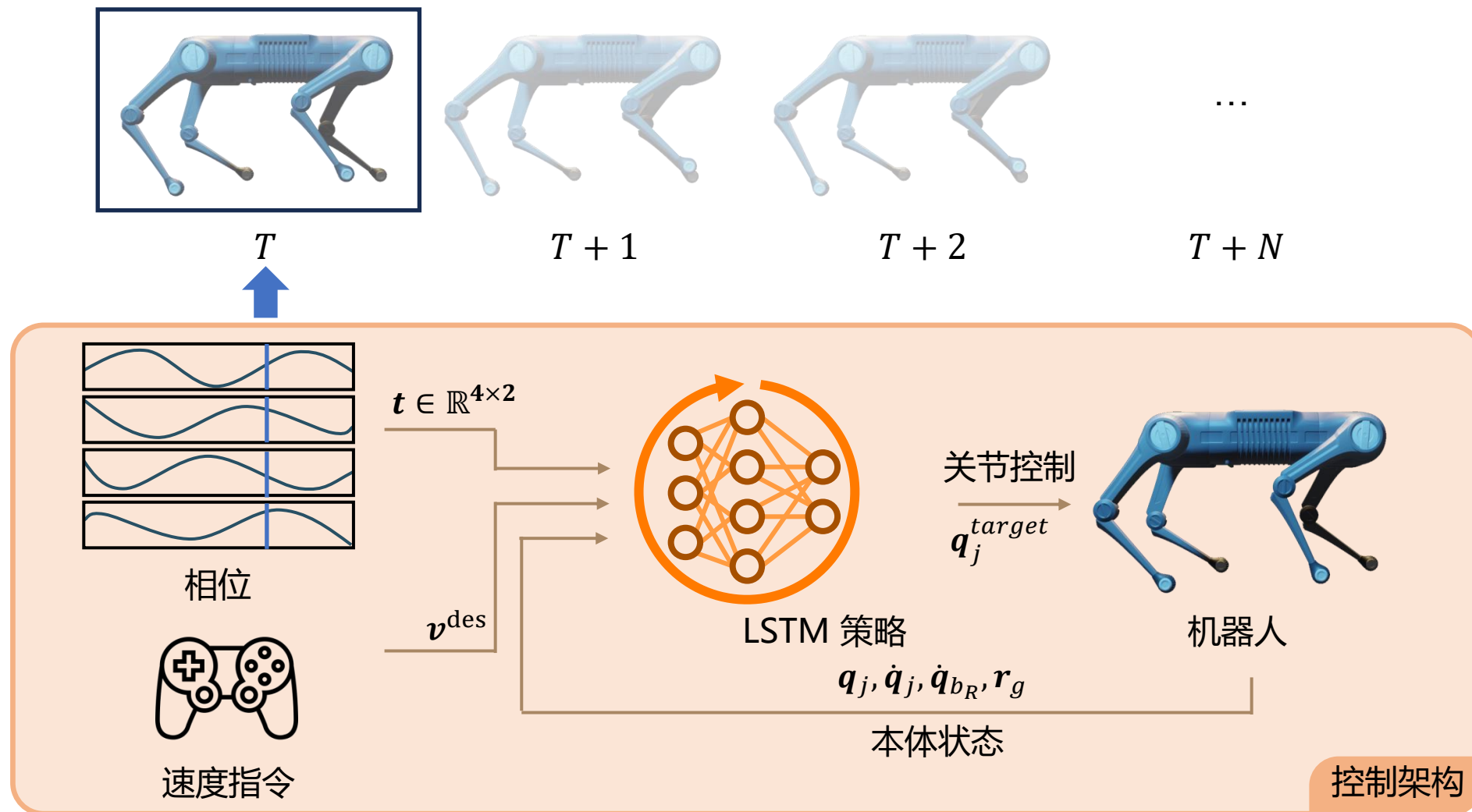
地形适应



- 使用**落脚点**作为引导信号
- 定量**描述地形**差异

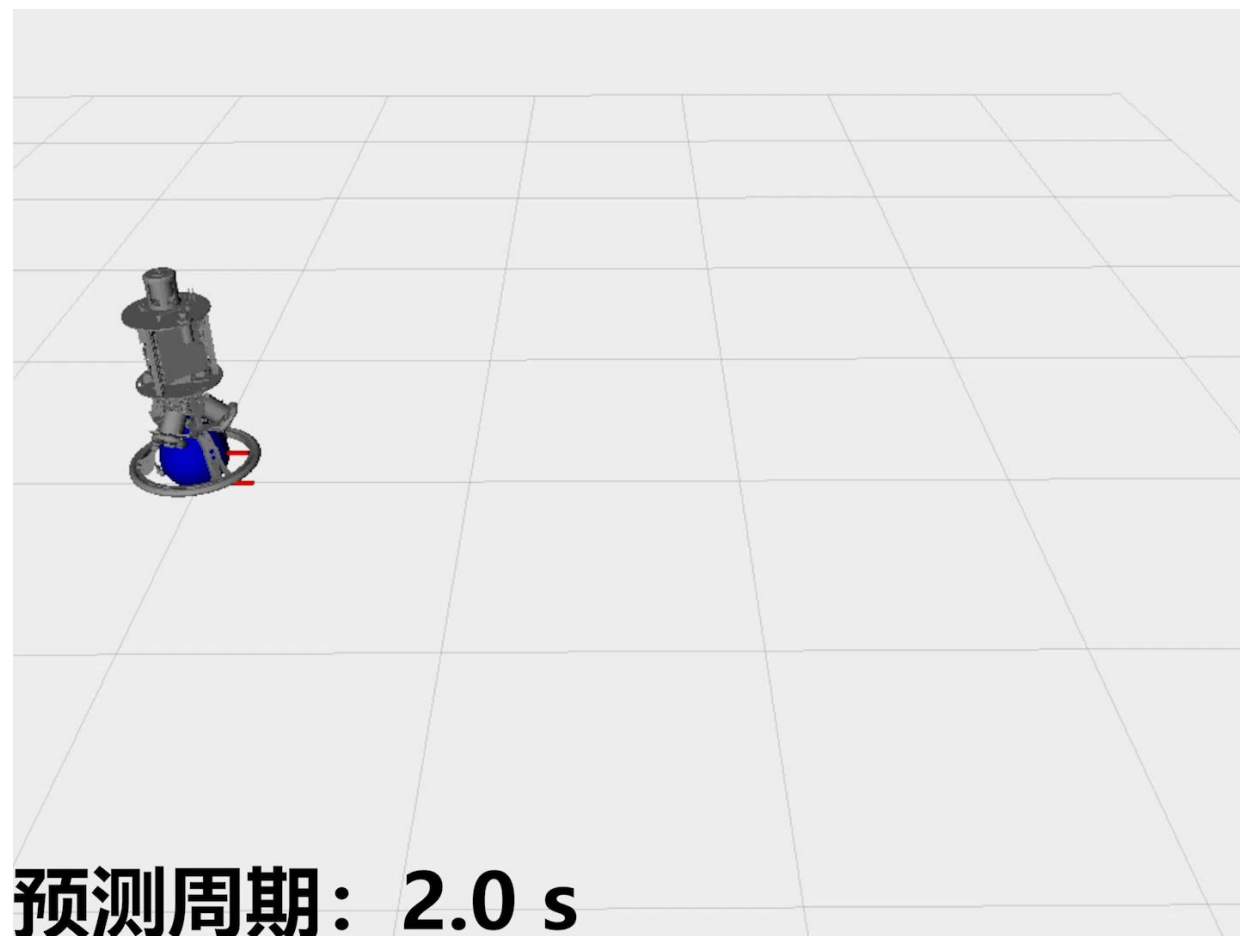
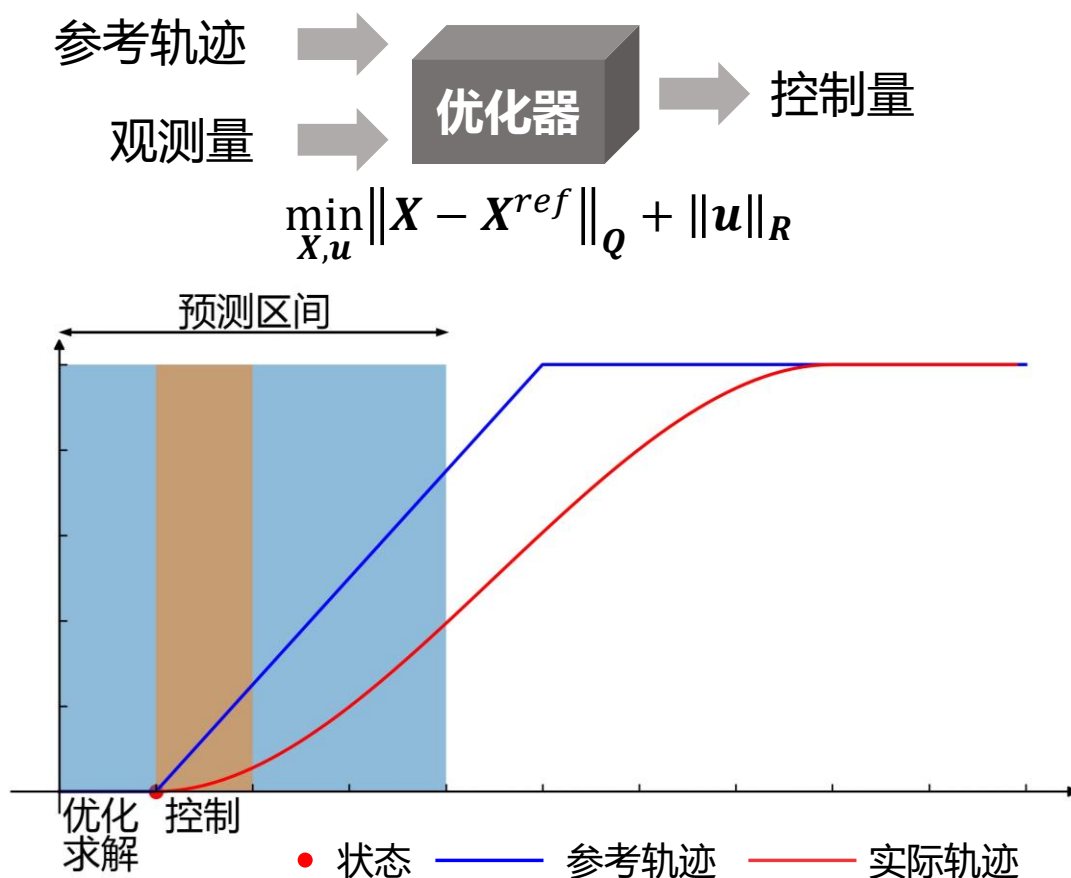


- 基于**相位**的编码方式只适用于**简单周期运动**，难以生成复杂动作
- **动作捕获**技术可以记录真实、复杂的动作，广泛应用于游戏与动画产业



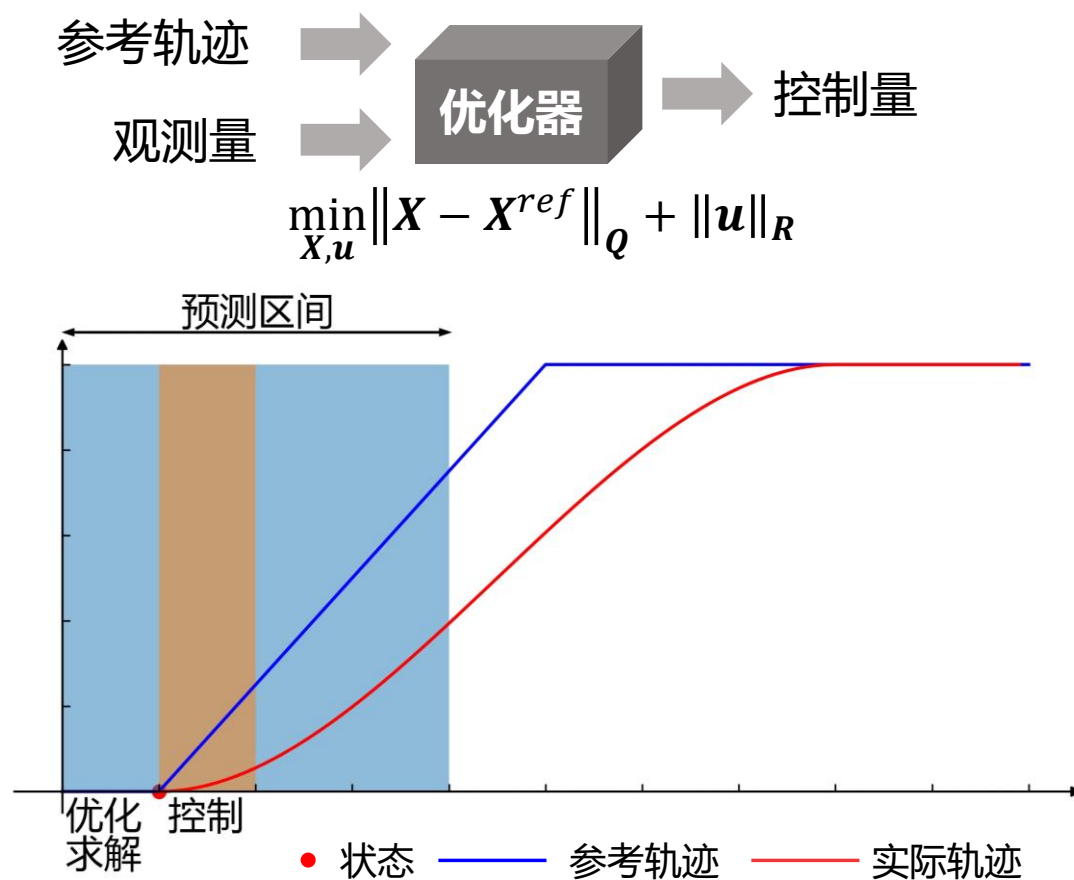
相位信号只能映射至**单步**的参考状态，无法直接提供**未来运动**信息

模型预测控制

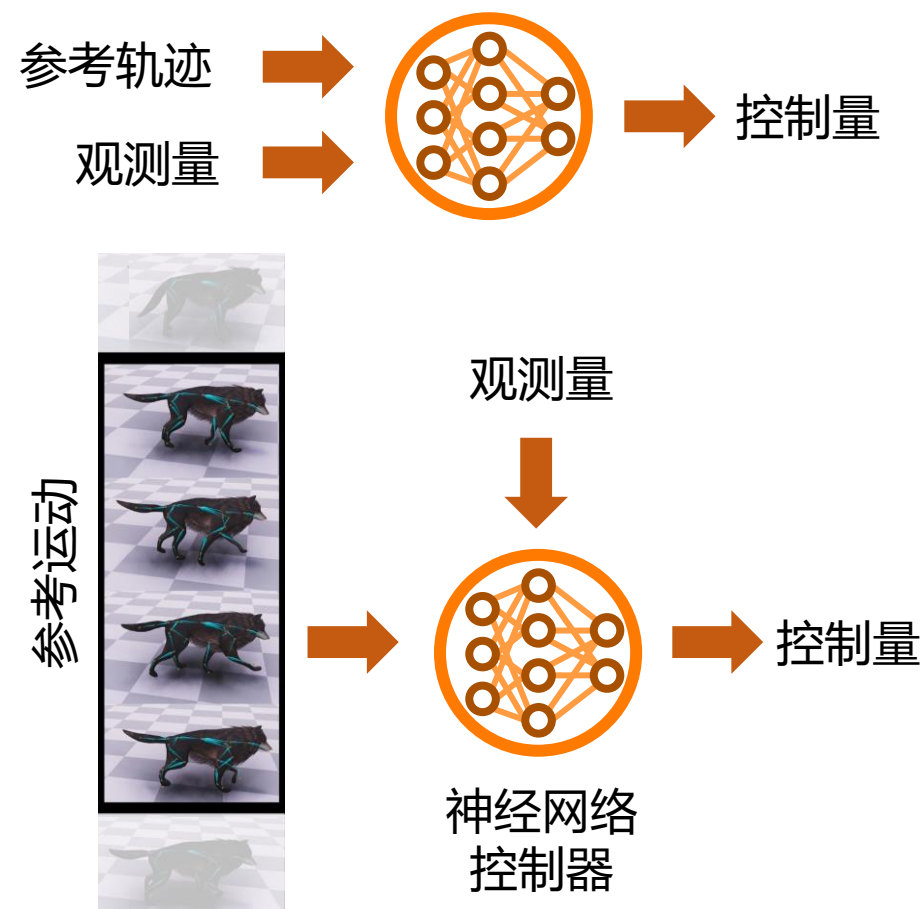


模型预测控制需要优化**未来**较长一段时间的**运动轨迹**，才能实现较好的控制效果

模型预测控制



动作模仿强化学习控制



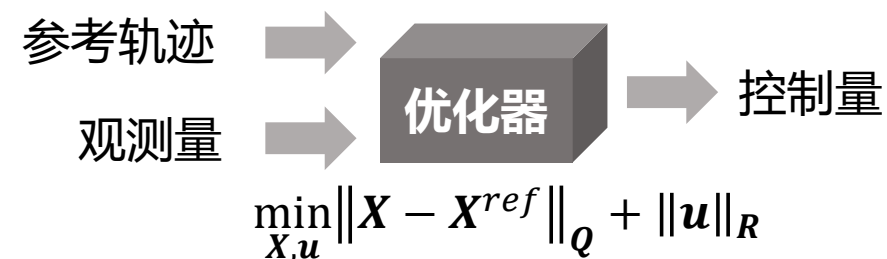
- 两种方法解决的**问题相同**，均根据**参考运动**与当前状态，求解控制量
- 基于强化学习的方法也受未来**参考轨迹长度**的影响

模型预测控制

□ **在线**优化方法本身**不包含**参考运动信息，需要在求解时提供**完整**的未来参考运动信息



完整参考运动



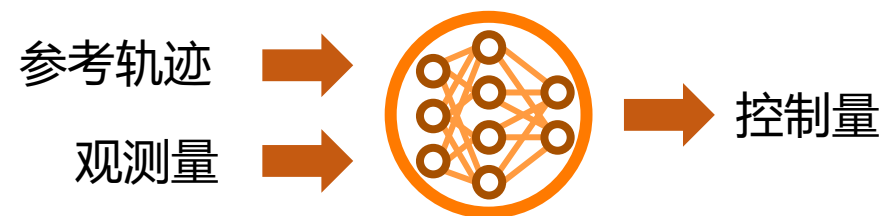
动作模仿强化学习控制

□ **离线**优化方法可在训练时**习得**部分参考运动信息，根据已知运动**预测补充**未来运动信息



已知运动

预测运动



模型**预测能力**的上限由参考轨迹的**内在关联性 (可预测性)** 决定

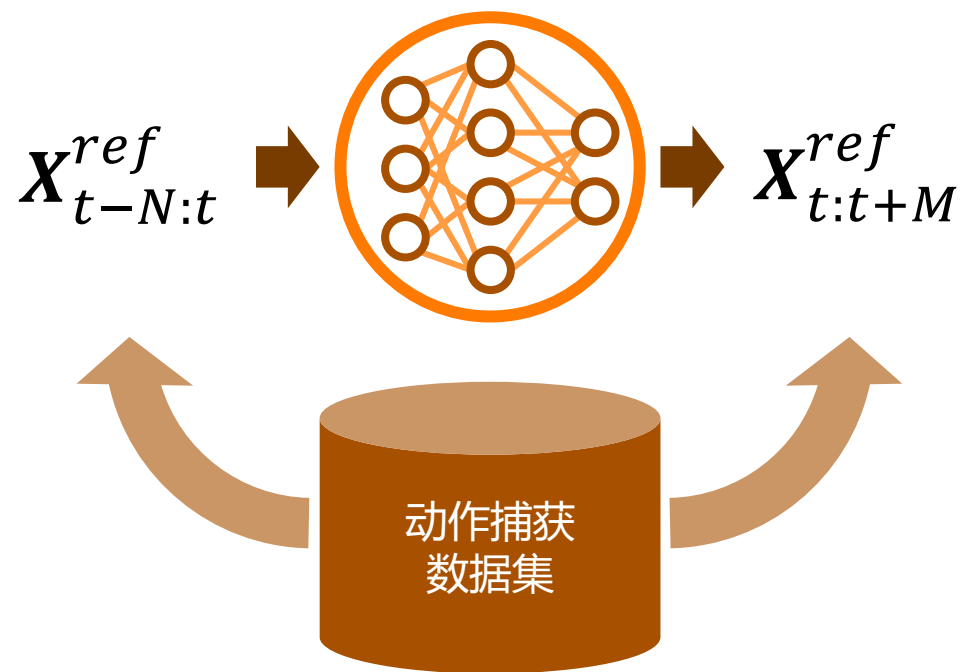
高维数据的关联性难以直接度量

通过神经网络拟合误差近似可预测性

$$\mathbf{X}_{t-N:t}^{ref} \in \mathbb{R}^{19N}$$

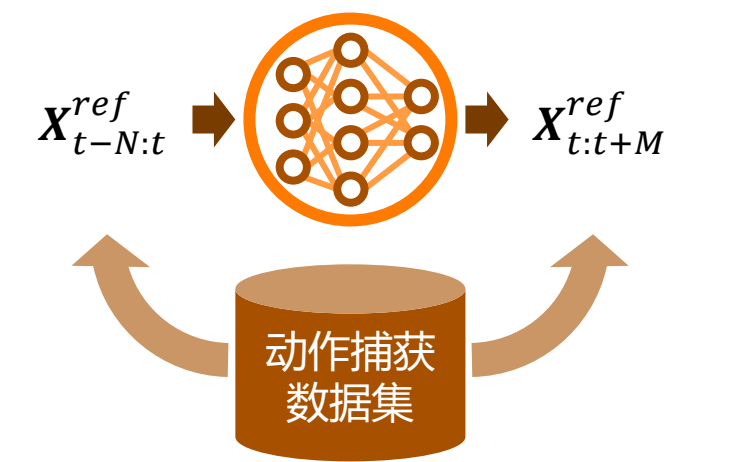
$$\mathbf{X}_{t:t+M}^{ref} \in \mathbb{R}^{19M}$$

$$\text{Corr}(\mathbf{X}_{t-N:t}^{ref}, \mathbf{X}_{t:t+M}^{ref}) = ?$$

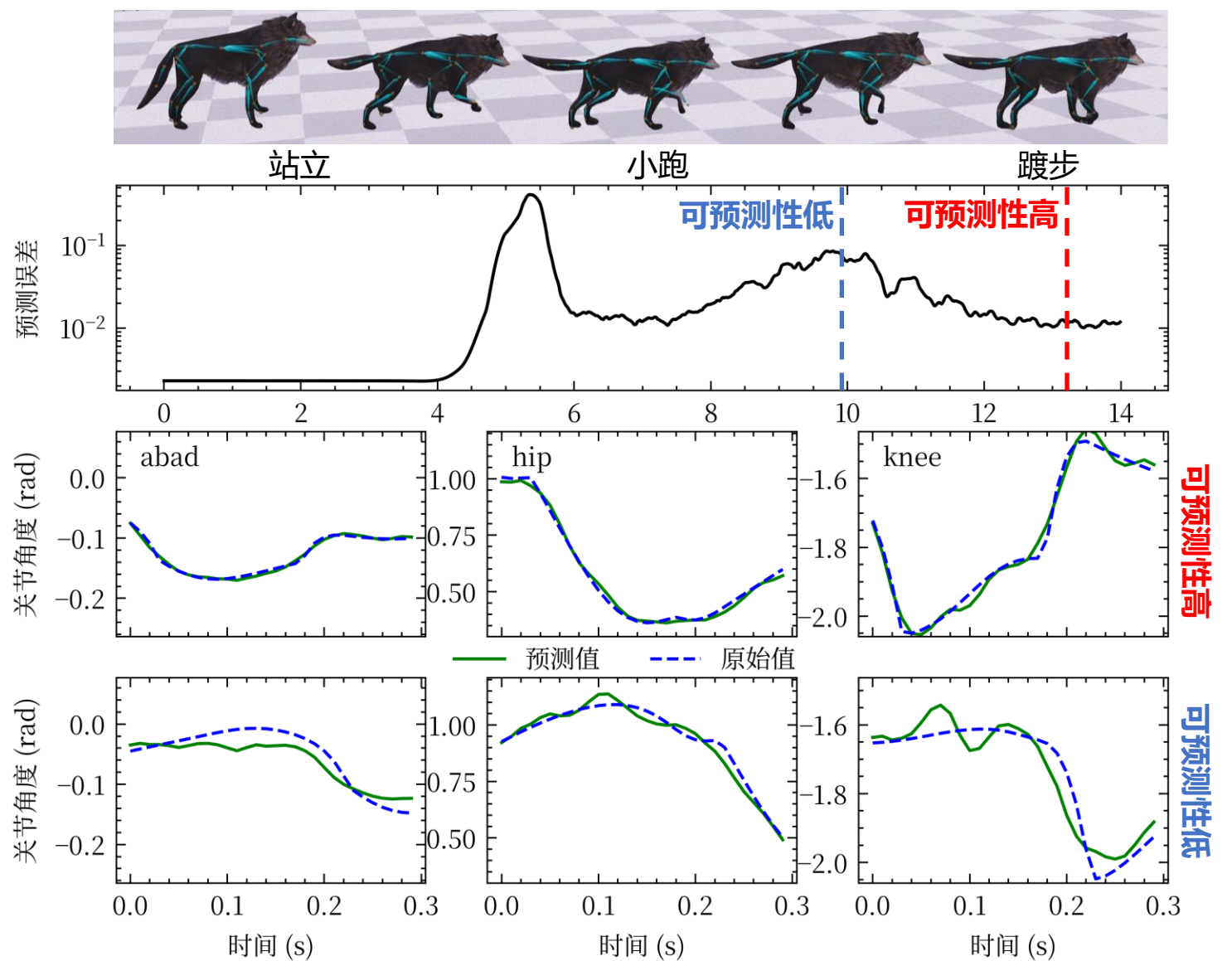


$$\text{预测误差}(t) = \left\| \mathbf{x}_{t:t+M}^{ref} - \hat{\mathbf{x}}_{t:t+M}^{ref} \right\|$$

参考轨迹可预测性度量

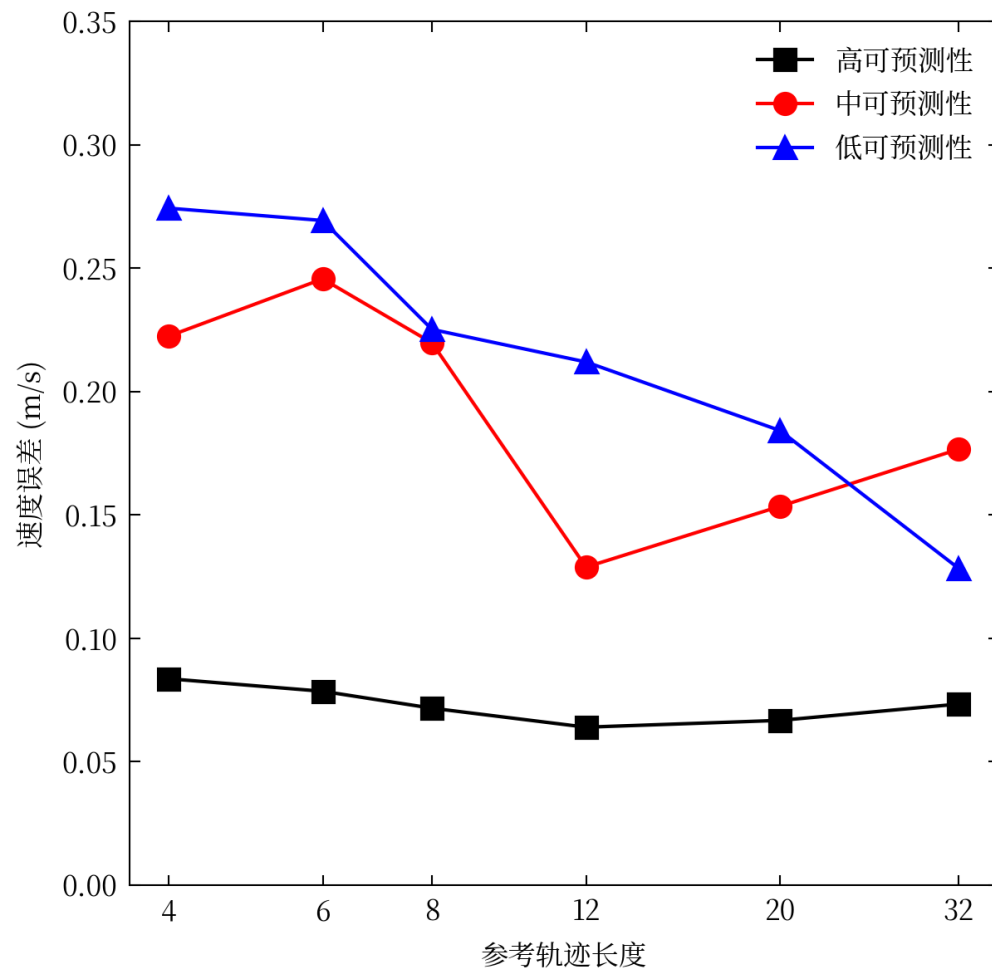


预测误差(t) = $\| \mathbf{x}^{ref}_{t:t+M} - \hat{\mathbf{x}}^{ref}_{t:t+M} \|$

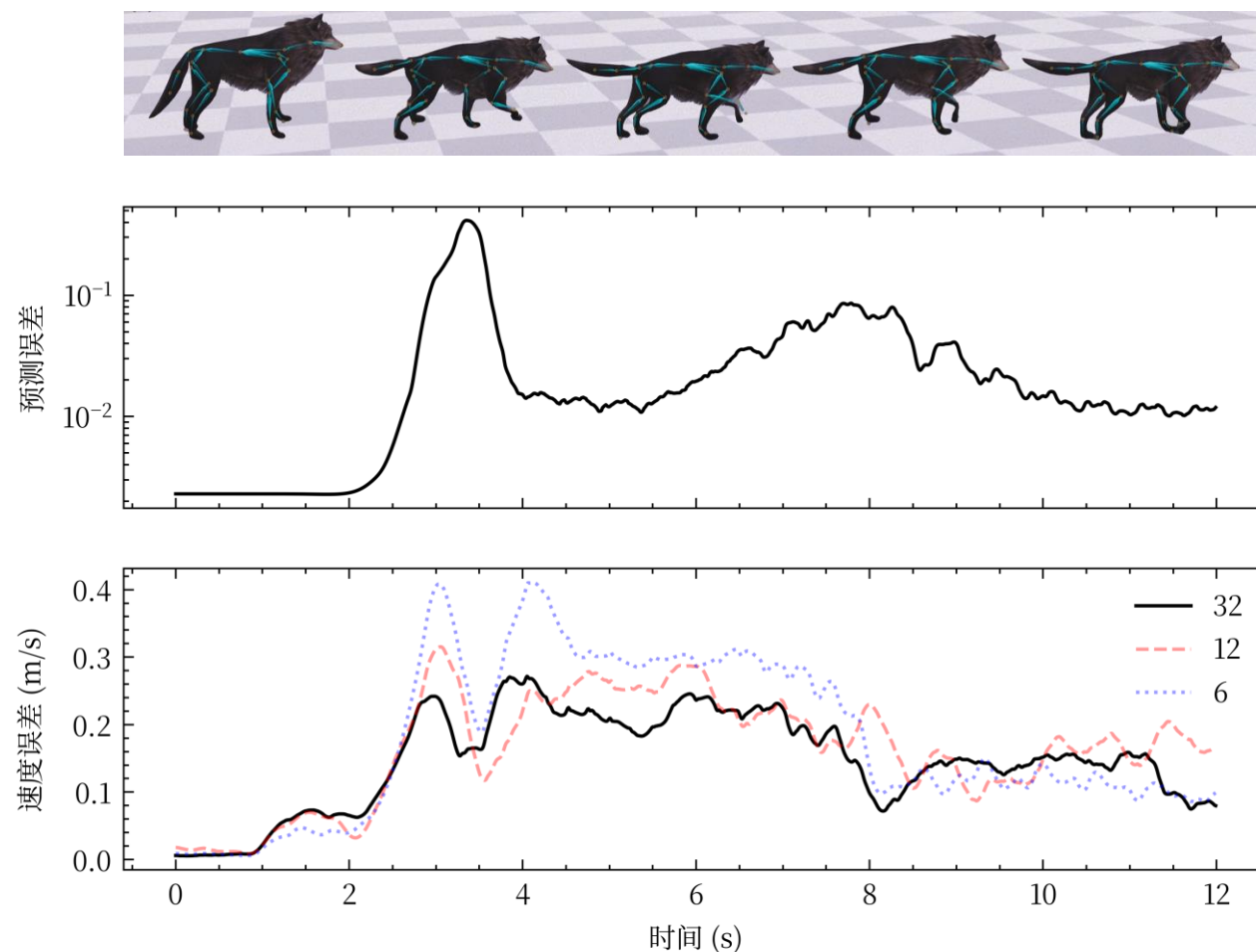


强化学习方法的**预测能力**由参考轨迹的**内在关联性 (可预测性)** 决定

统计结果

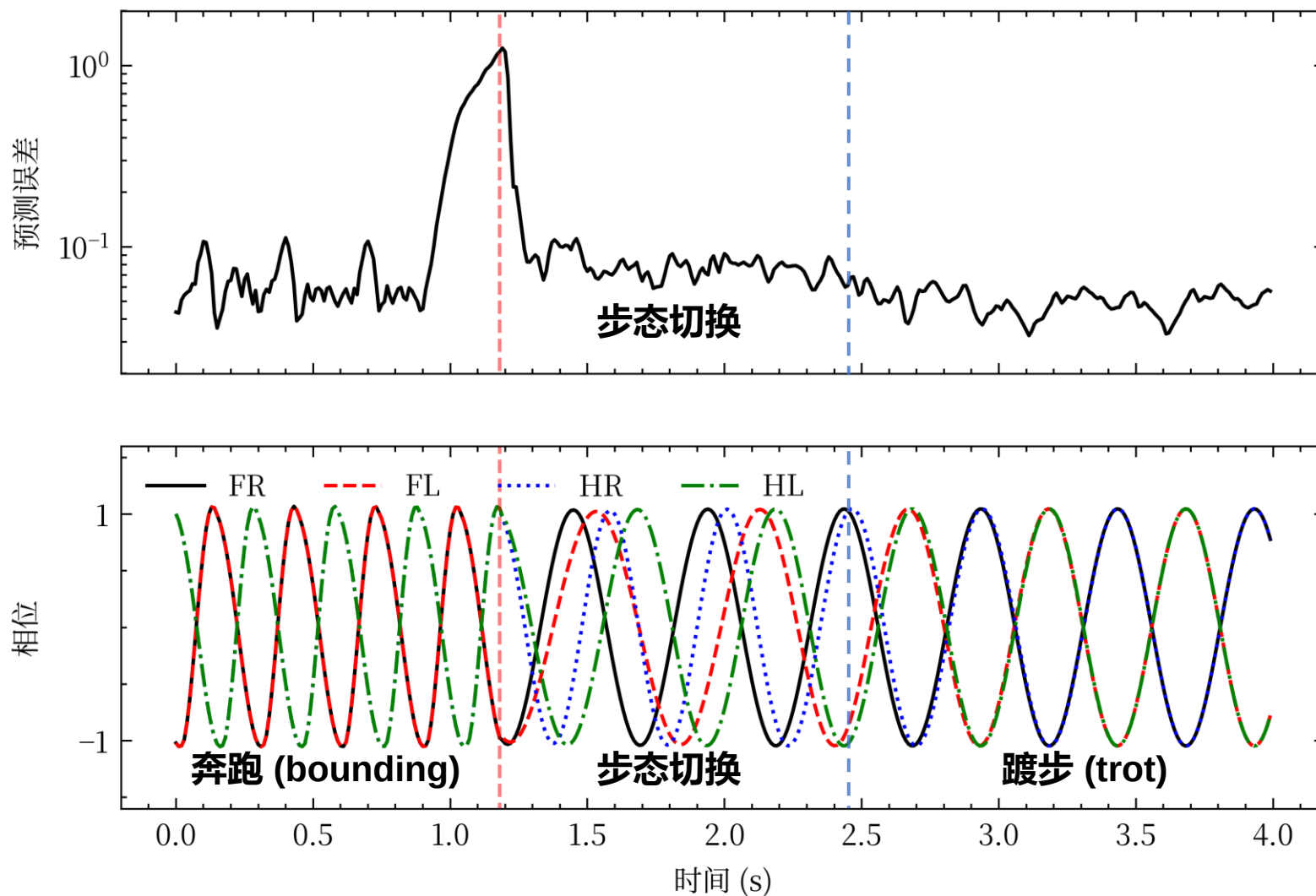


时域对比



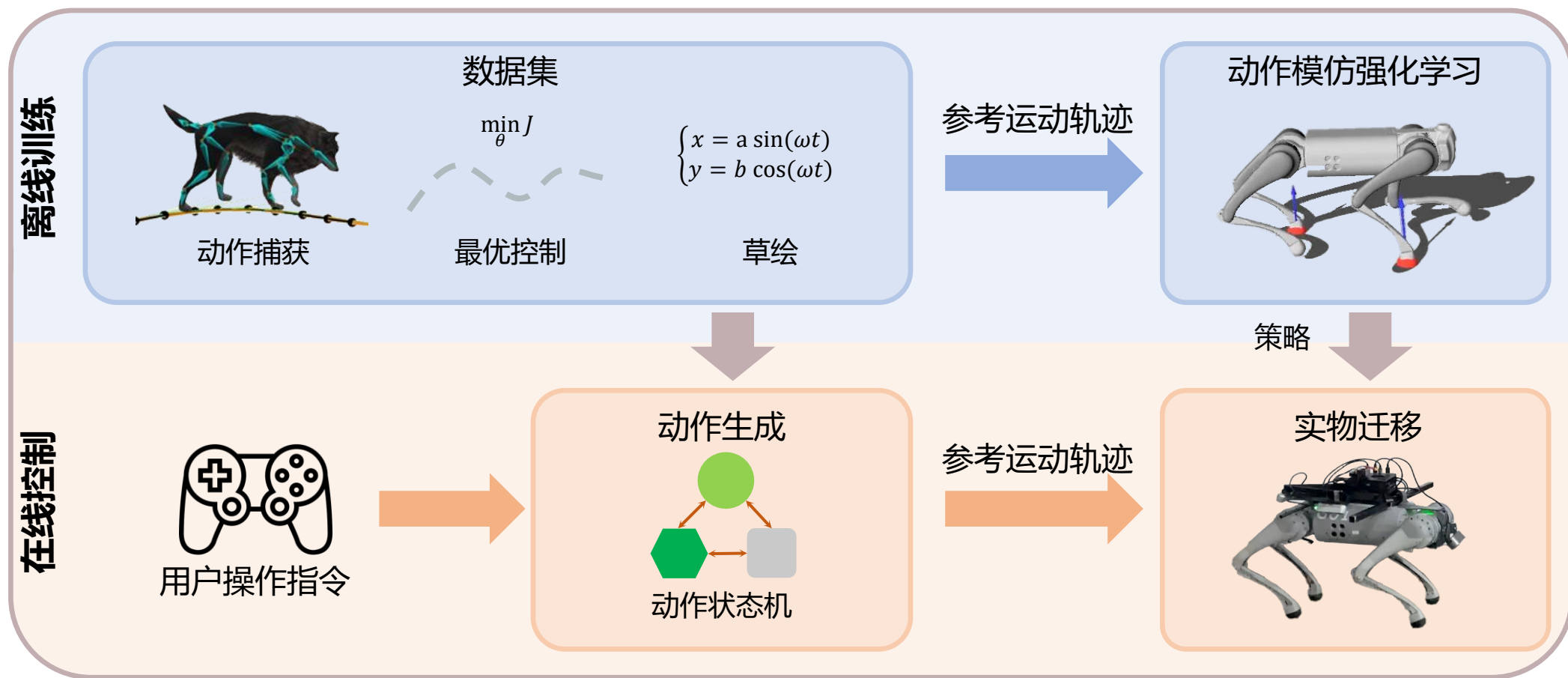
□ 参考轨迹相同时，可预测性**越高**，控制器的**性能越好**

□ 对于可预测性**较低**的动作，**延长**参考轨迹长度可以**提高**控制器的性能



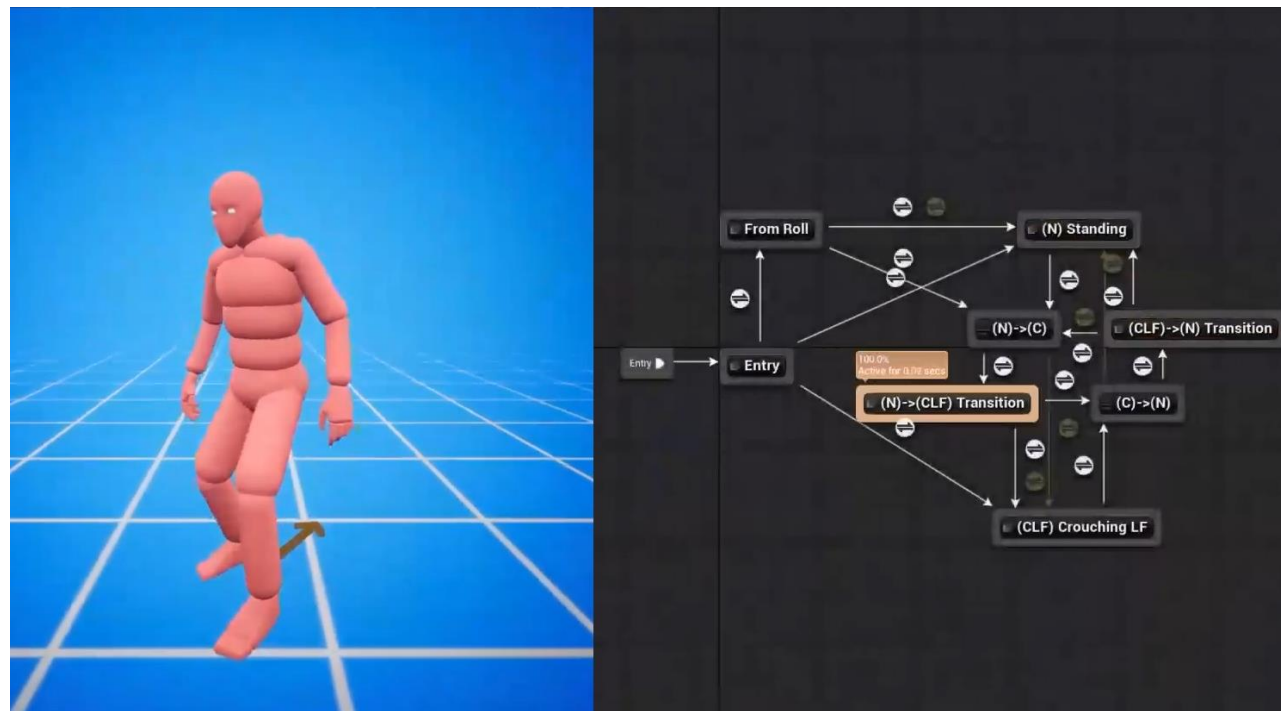
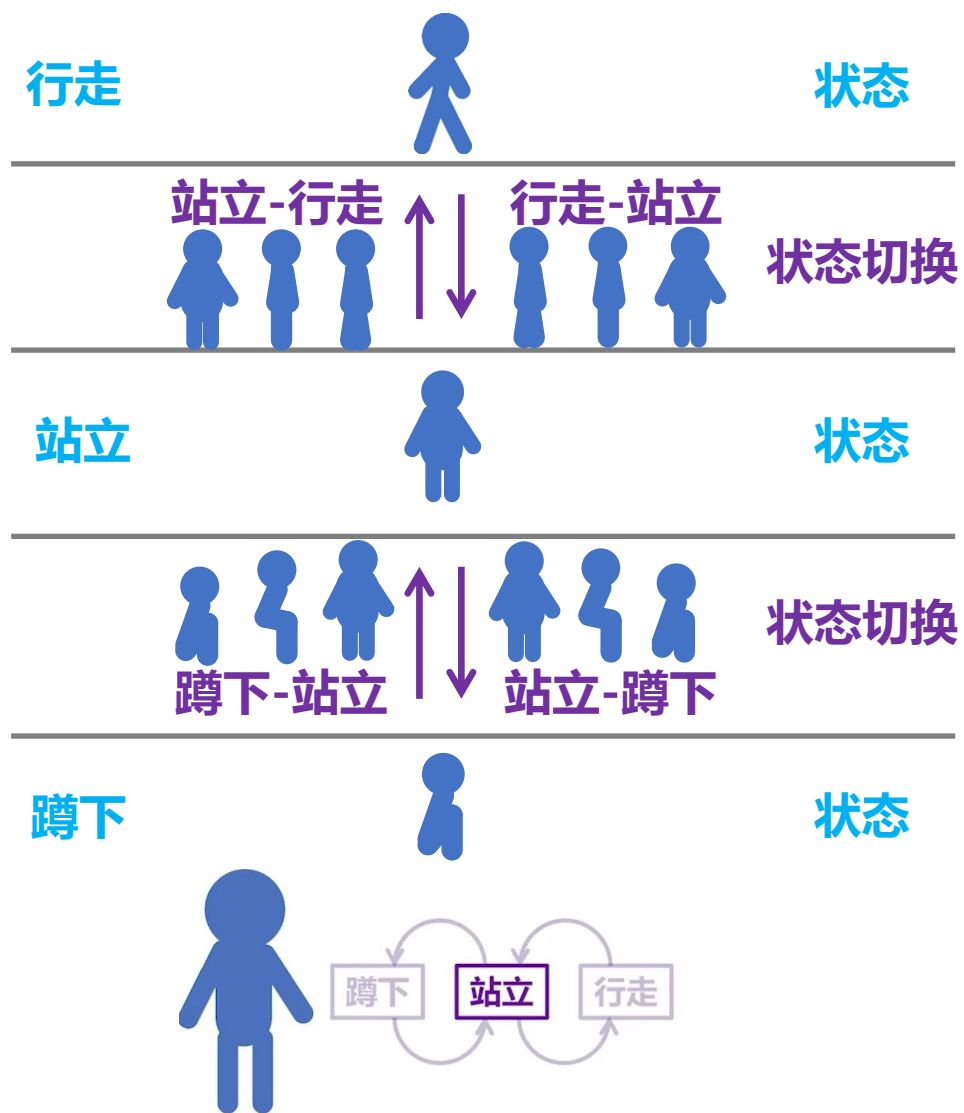
□ 相位生成的步态切换较为简单，与自然动物的步态切换不同

□ 步态切换过程中可预测性水平高，不需要更长的参考轨迹



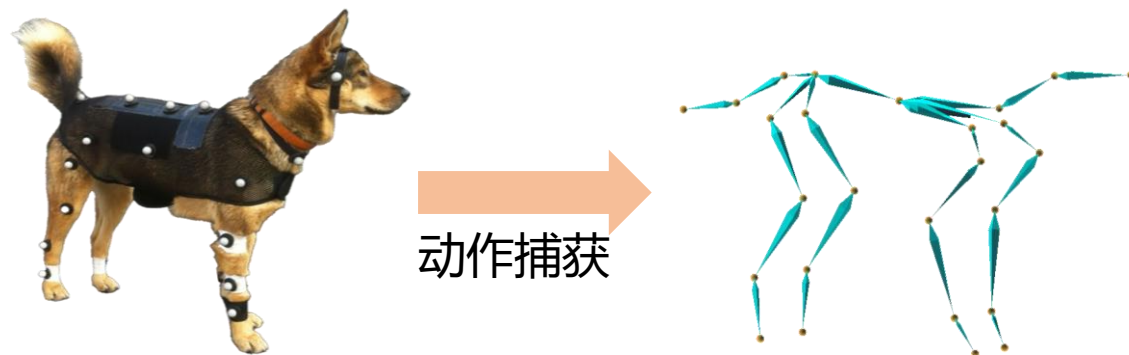
- 使用**参考运动轨迹**作为输入，**突破**了简单引导信号的**局限性**
- **运动学轨迹**引导同时适用于动作捕获、轨迹优化、运动学草绘等**多种来源**的动作
- 在控制中，使用**动作状态机**根据用户指令，将固定动作片段**合成**连贯可控动作

有限状态机是目前游戏领域发展**最成熟**、应用**最广泛**的动作生成技术

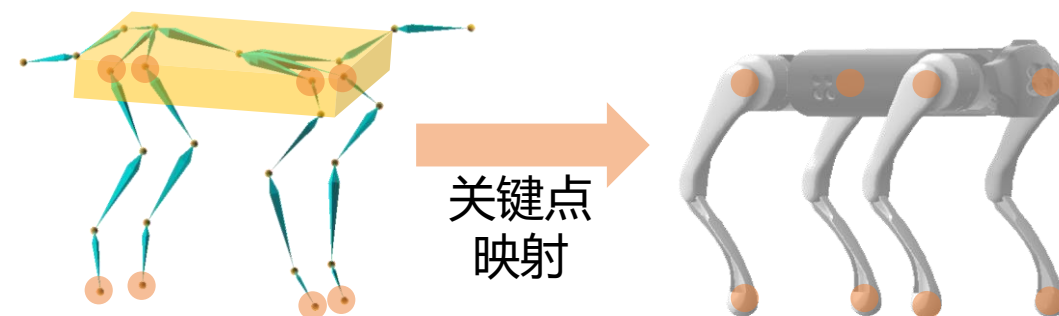


不同状态之间的排列组合生成**连贯**、**可控**的复杂动作

动作捕获：公开数据库



动作重映射：消除构型差异



全向运动状态机

高速 左转	高速 前进	高速 右转	小跑 trot
低速 左转	低速 前进	低速 右转	踱步 pace
原地 左转	原地 站立	原地 右转	站立 stand

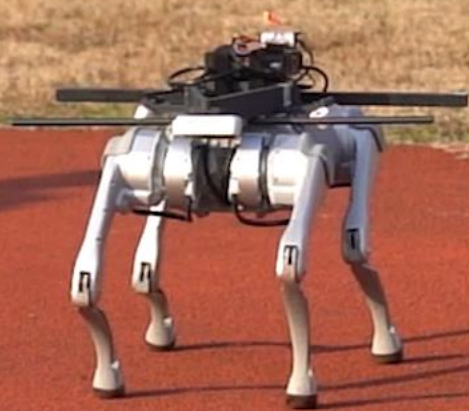
- 任意两种状态之间均可进行切换
- 状态切换的数据中包含了**天然**的步态切换

动作捕获运动重现

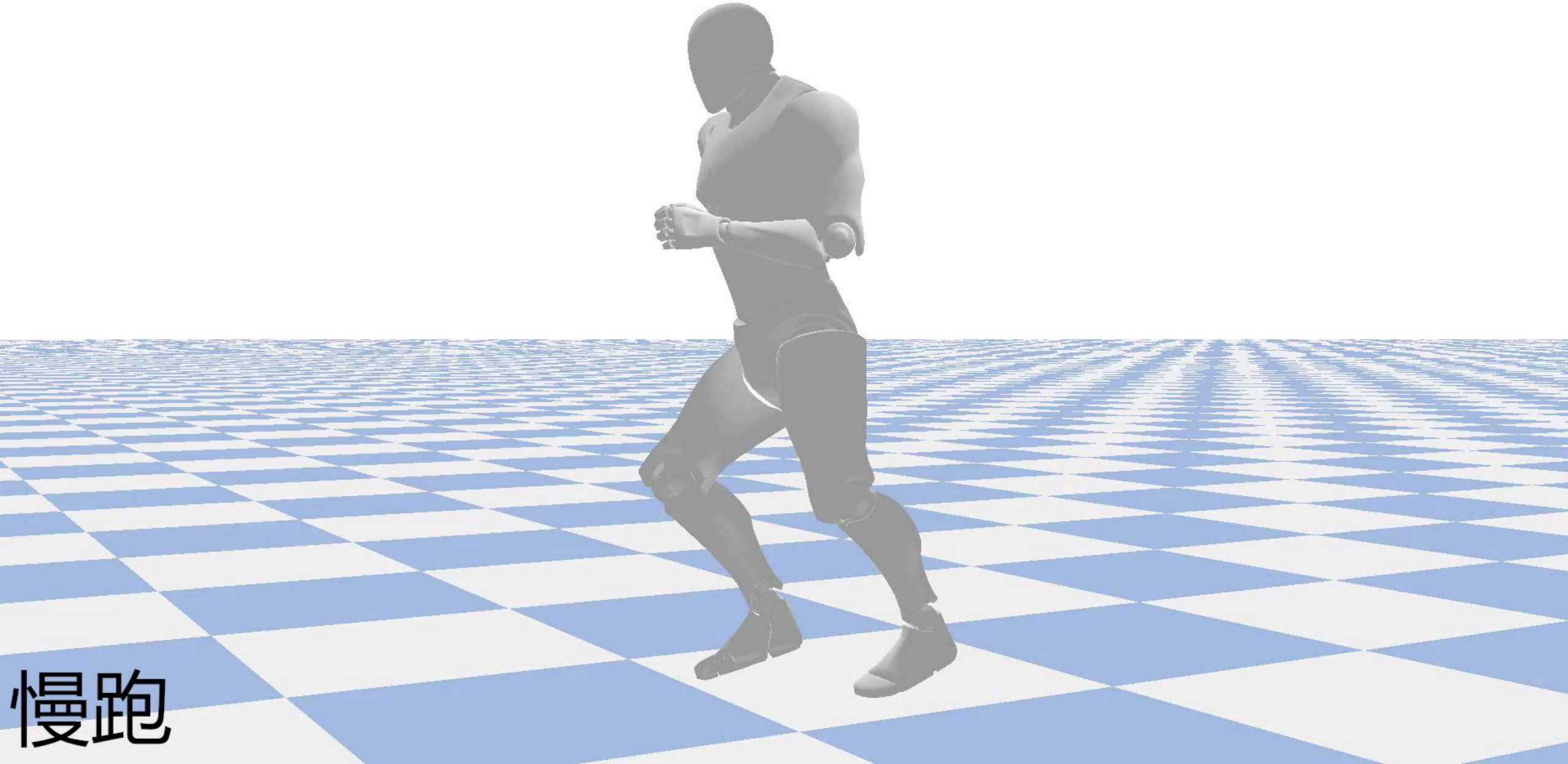


控制器在机器人上重现**动作捕获**中动物自然、真实的步态切换过程

基于状态机的指令跟随

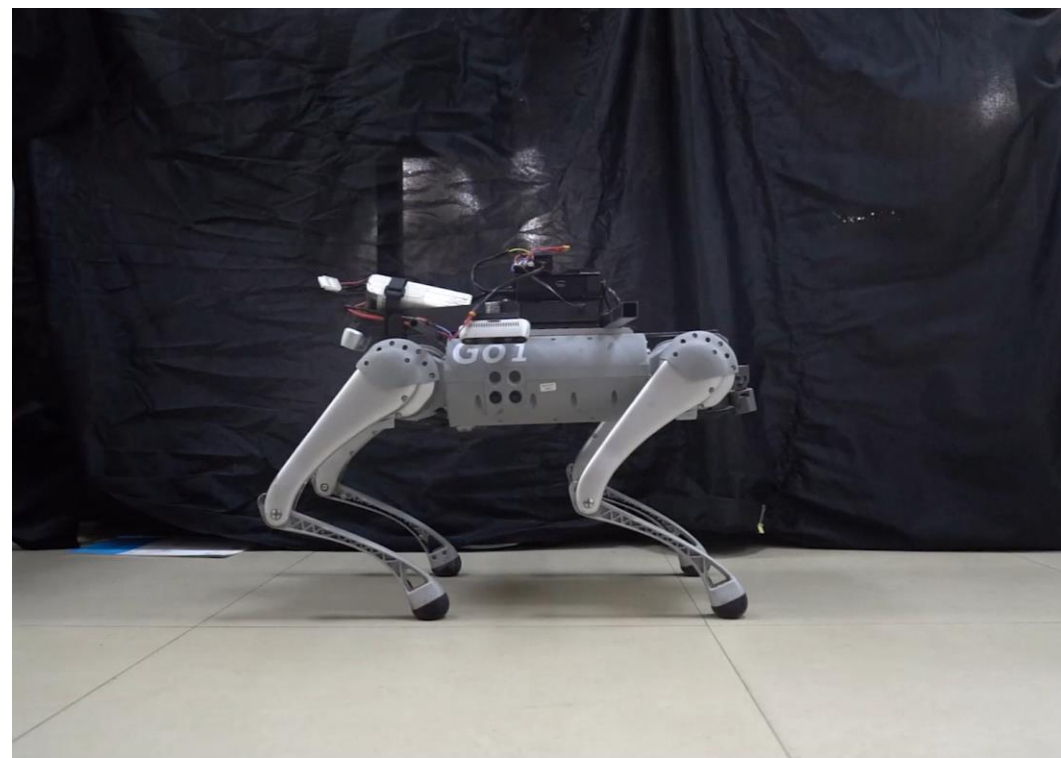
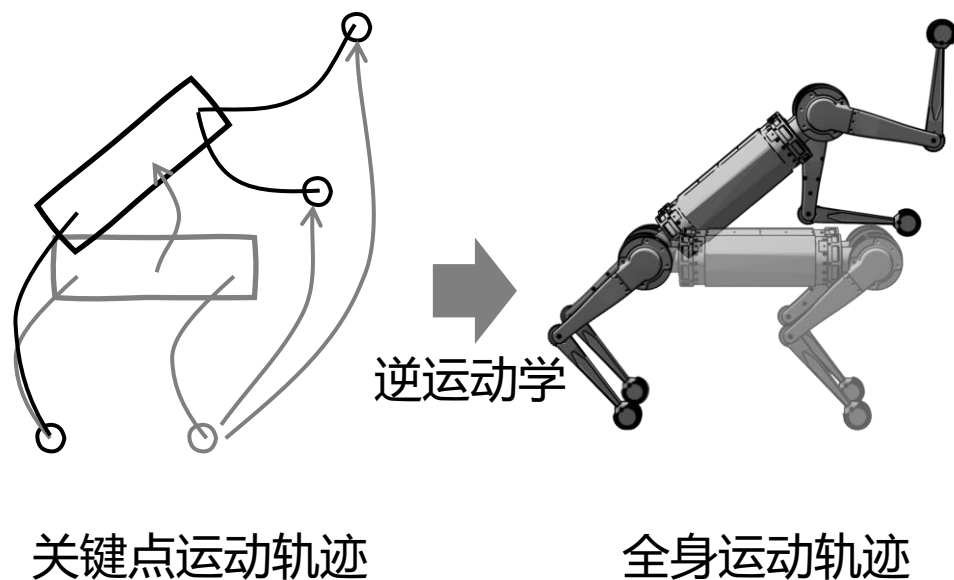


(预研) 双足机器人运动控制



慢跑

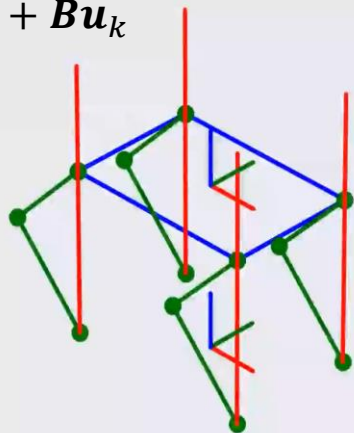
运动学草绘



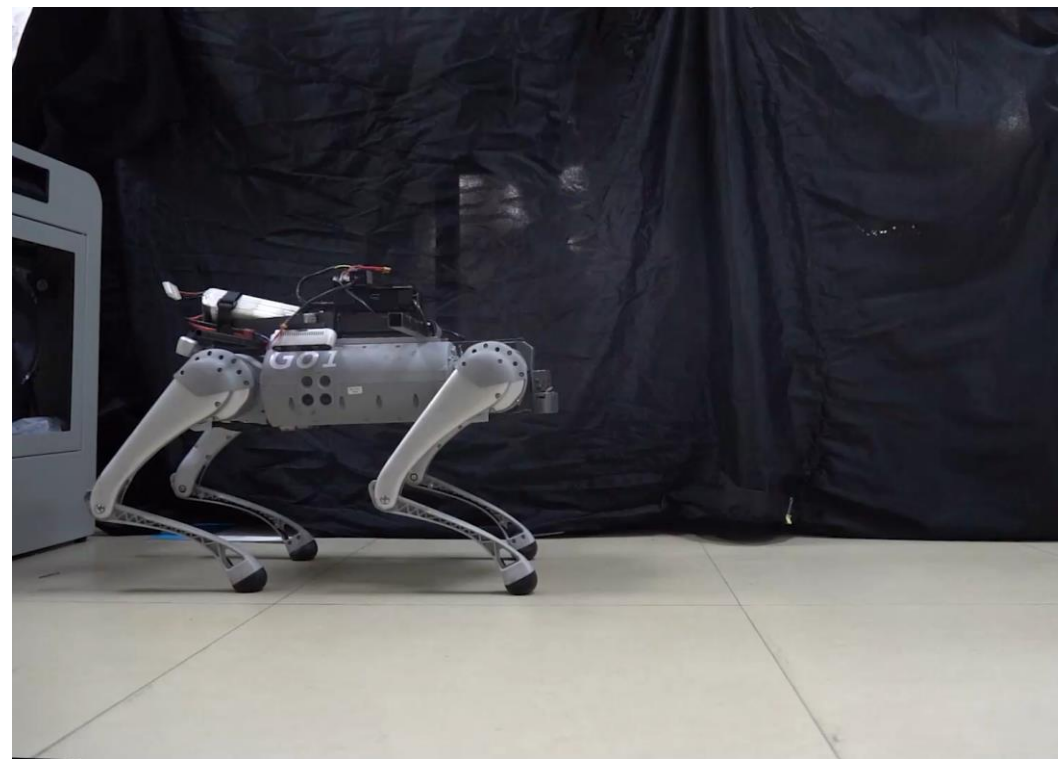
- 根据任务目标草绘**身体与足端**运动轨迹，通过逆运动学获得**全身**运动轨迹
- 运动学草绘**易于**获取符合预期的动作

轨迹优化

$$\begin{aligned} \min_{X,u} & \|X - X^{ref}\|_Q + \|u\|_R \\ \text{s.t. } & X_{k+1} = AX_k + Bu_k \end{aligned}$$



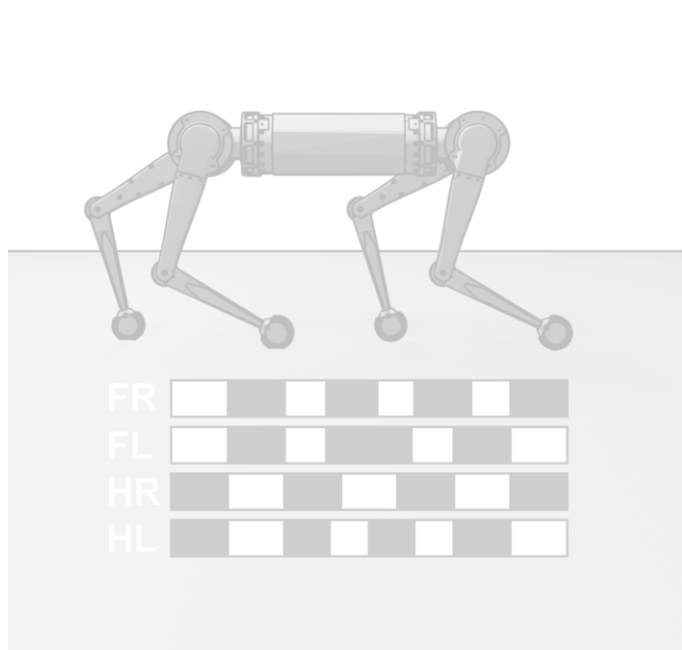
草绘



轨迹优化可获取更接近**动力学可行域**的参考运动轨迹

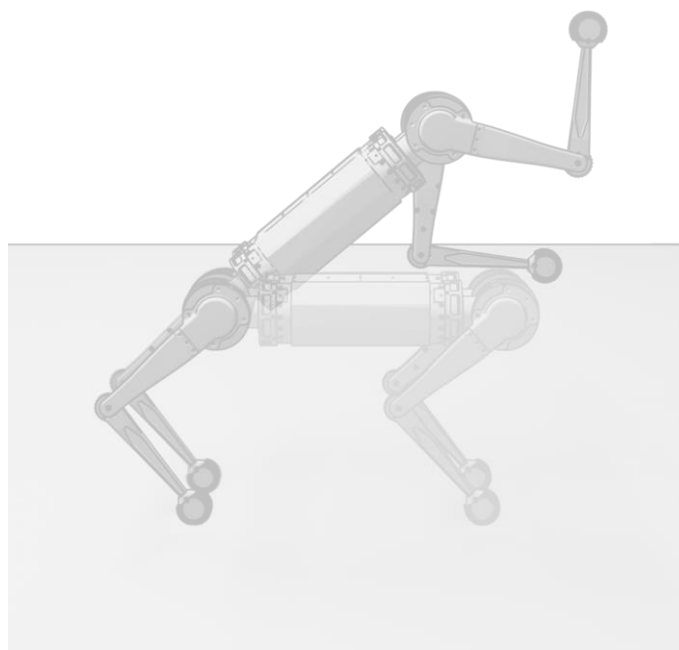
多任务学习

步态切换



- 基于**相位**引导的控制器
- 分析**引导信号**的作用

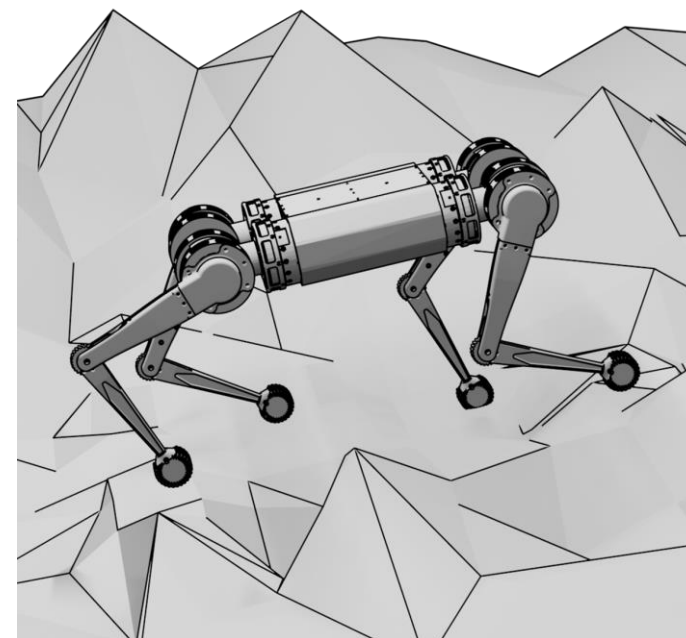
复杂动作



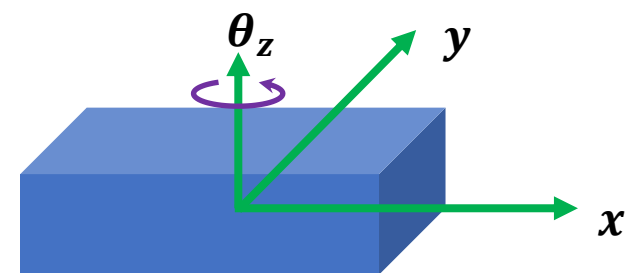
- 使用**运动学**轨迹引导
- 探索参考轨迹**时间关联**

泛化迁移

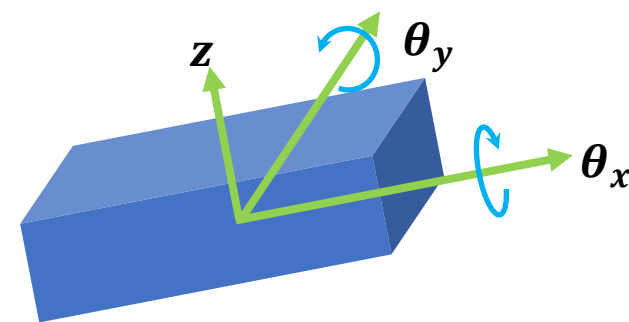
地形适应



- 使用**落脚点**作为引导信号
- 定量**描述地形**差异

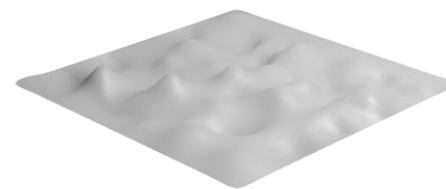


平地运动主要自由度

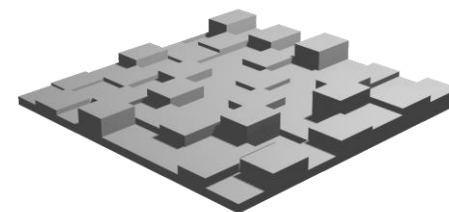


复杂地形运动额外自由度

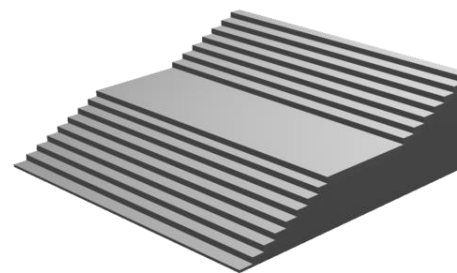
复杂地形运动中，机器人的**身体高度**、**姿态**随地形变化，动力学过程也更加复杂



连续起伏地形



离散起伏地形



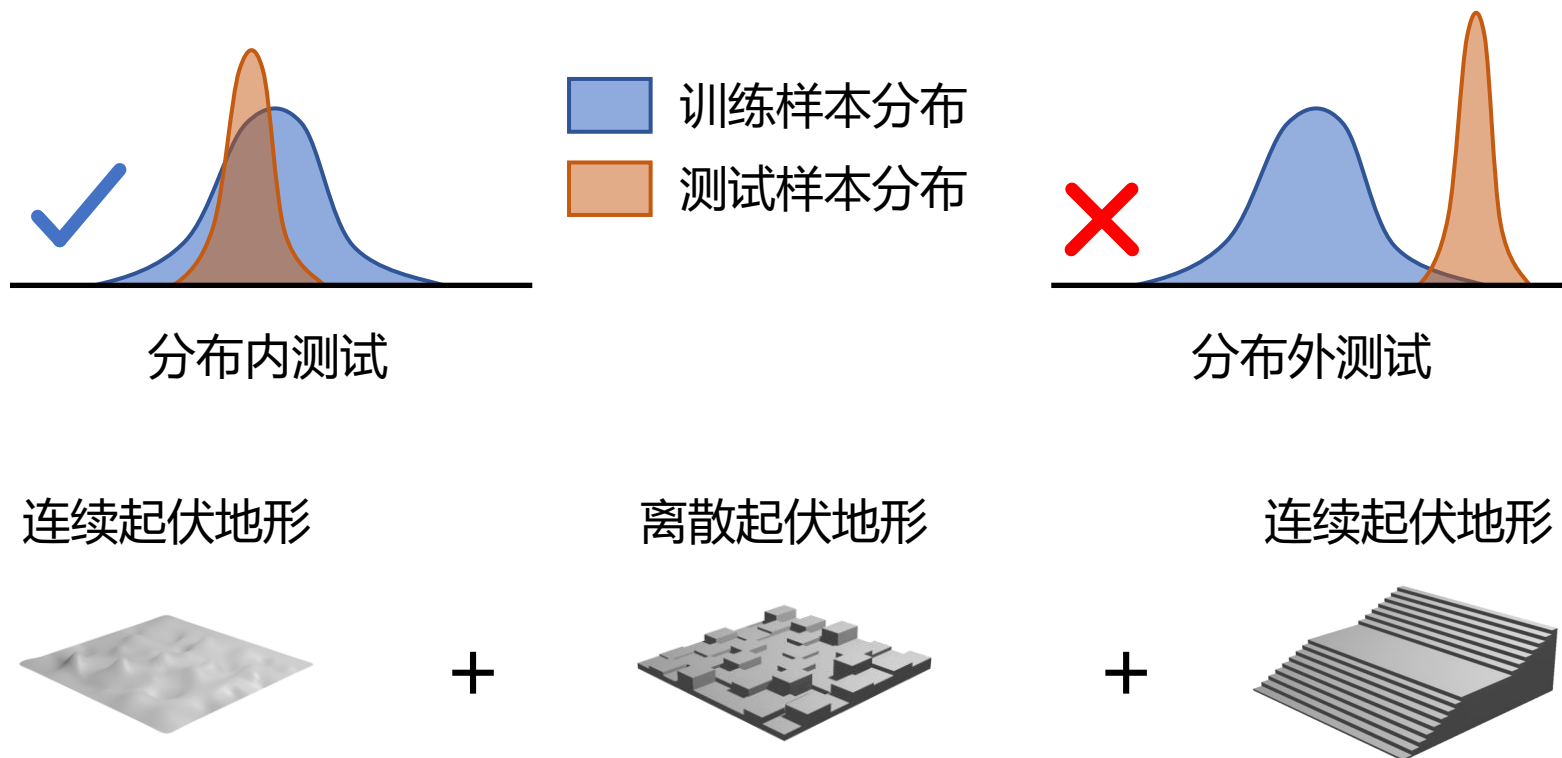
连续起伏地形

...

其他地形...

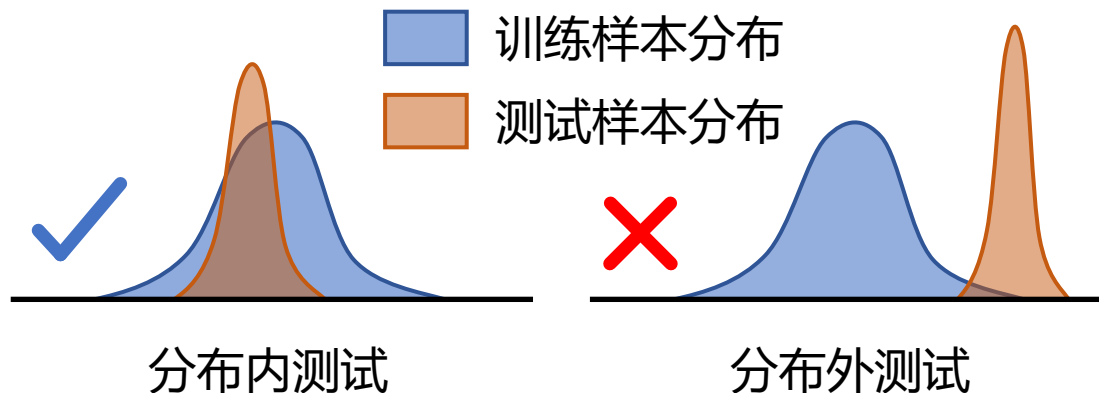
- 对于**不同种类地形**，控制算法需要应对的**动力学过程不同**
- 运动控制中的**地形适应问题**属于机器学习中的**泛化能力问题**

机器学习模型的泛化能力由**训练样本分布**决定，一般样本分布越广泛，泛化能力越强



经验方法：使用常用地形的组合扩大训练样本分布

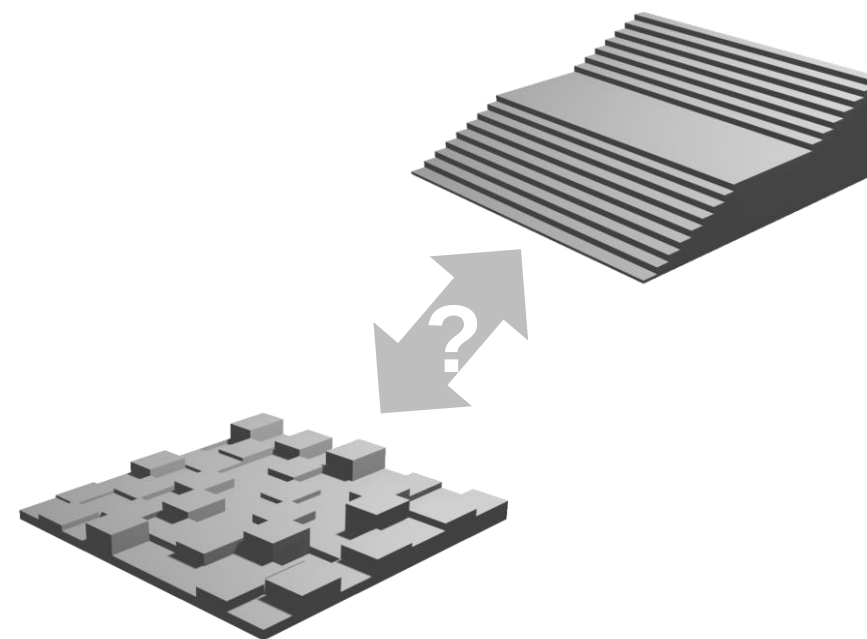
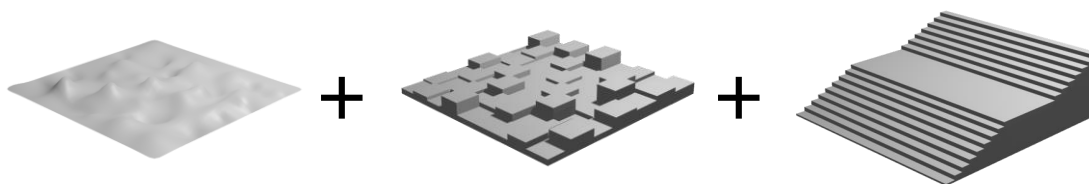
机器学习模型的泛化能力由**训练样本分布**决定，一般样本分布越广泛，泛化能力越强



连续起伏地形

离散起伏地形

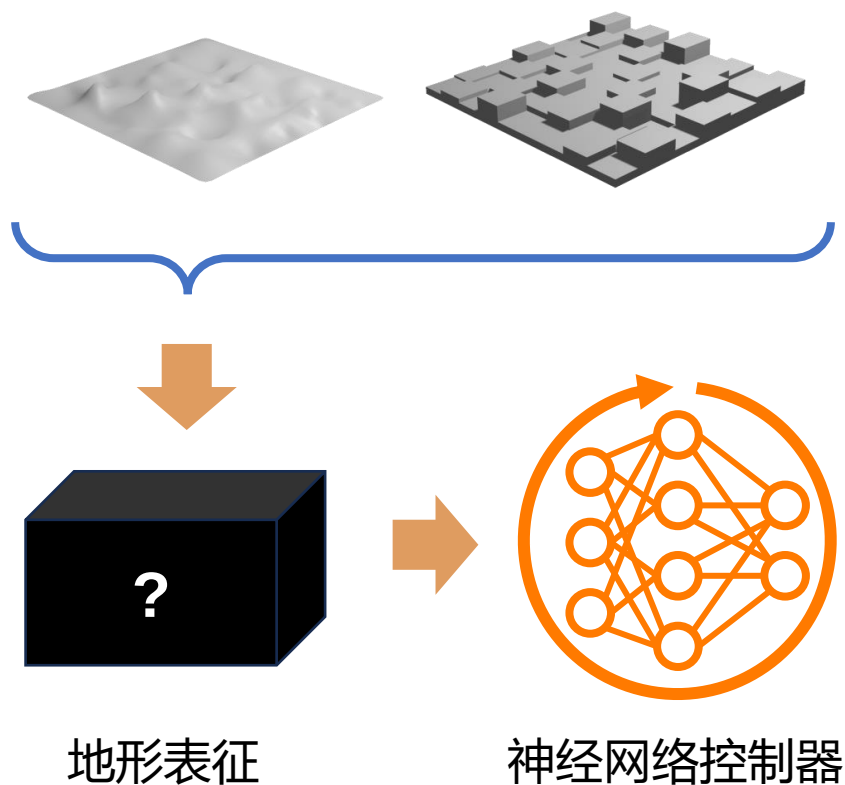
连续起伏地形



经验方法：使用常用地形的组合扩大训练样本分布

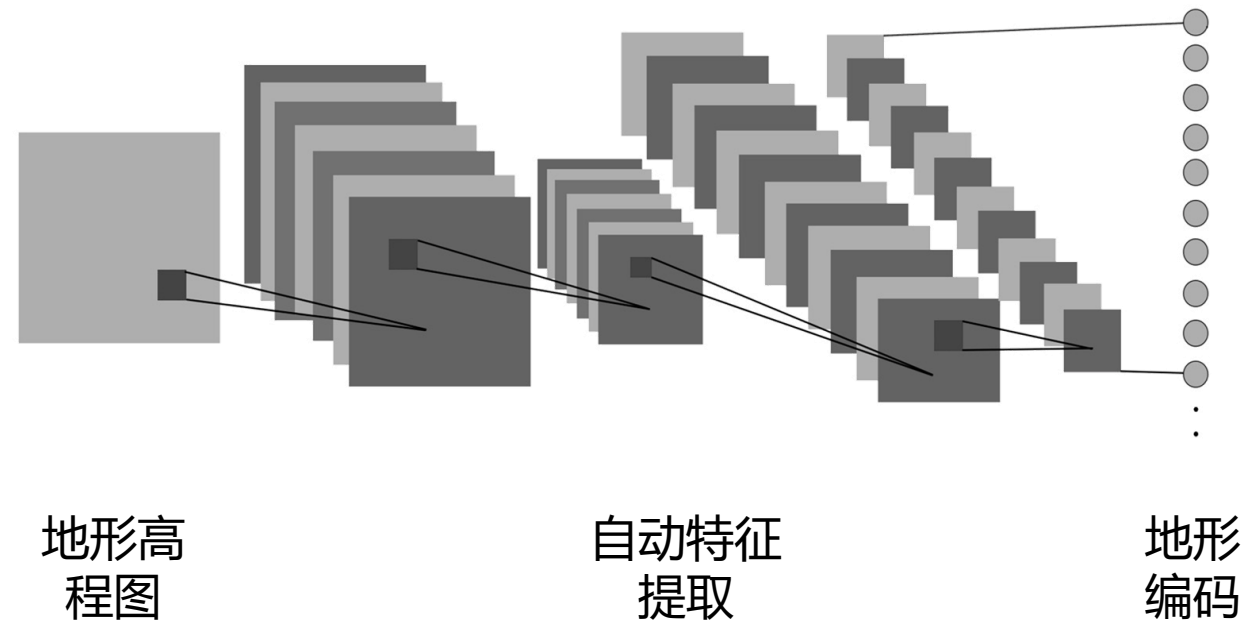
如何**度量**不同地形下样本**分布**的差异？

控制器设计的难点

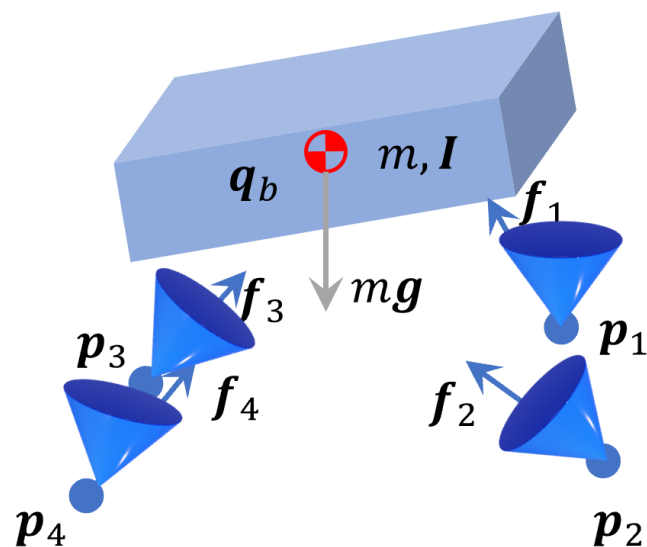


如何**高效**地将地形信息**编码**为控制器输入

卷积神经网络



- ❑ 基于卷积神经网络进行编码的方式**计算效率低**
- ❑ 难以直观理解**地形与运动控制**的关系



单刚体动力学模型

$$\ddot{q}_{b_L} = \frac{1}{m} \sum_{i=1}^N f_i + g$$

$$\ddot{q}_{b_R} = I^{-1} \sum_{i=1}^N (q_{b_L} - p_i) \times f_i$$

优化变量

$$q_b(t) \in \mathbb{R}^6, p_i(t) \in \mathbb{R}^3, f_i(t) \in \mathbb{R}^3, c_i(t) \in \{0,1\}$$

约束条件

$$\ddot{q}_b = \mathcal{D}(q_b, p_{i=1,\dots,N}, f_{i=1,\dots,N}) \quad \text{动力学模型}$$

$$p_i \in \mathcal{K}(q_b) \quad \text{几何约束}$$

$$\text{支撑相} \quad c_i = 1$$

$$\text{摆动相} \quad c_i = 0$$

$$p_i^z = \mathcal{G}(p_i^{x,y})$$

$$p_i^z > \mathcal{G}(p_i^{x,y})$$

足端高度

$$\dot{p}_i = 0$$

足端速度

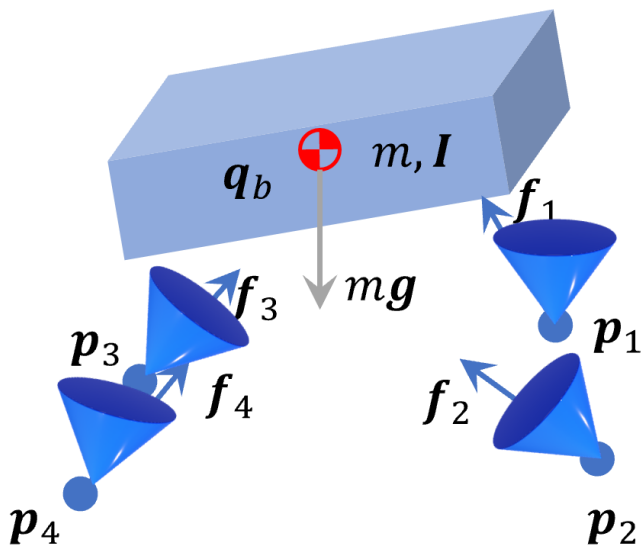
$$f_i \cdot n(p_i^{x,y}) > 0$$

$$f_i \in \mathcal{F}(\mu, n)$$

$$f_i = 0$$

足端力

$$f_i \in \mathcal{A}(q_b, p_i)$$



单刚体动力学模型

$$\ddot{q}_{b_L} = \frac{1}{m} \sum_{i=1}^N f_i + g$$

$$\ddot{q}_{b_R} = I^{-1} \sum_{i=1}^N (q_{b_L} - p_i) \times f_i$$

优化变量

$$q_b(t) \in \mathbb{R}^6, p_i(t) \in \mathbb{R}^3, f_i(t) \in \mathbb{R}^3, \mathcal{C}_i(t) \in \{0,1\}$$

约束条件

$\ddot{q}_b = \mathcal{D}(q_b, p_{i=1,\dots,N}, f_{i=1,\dots,N})$

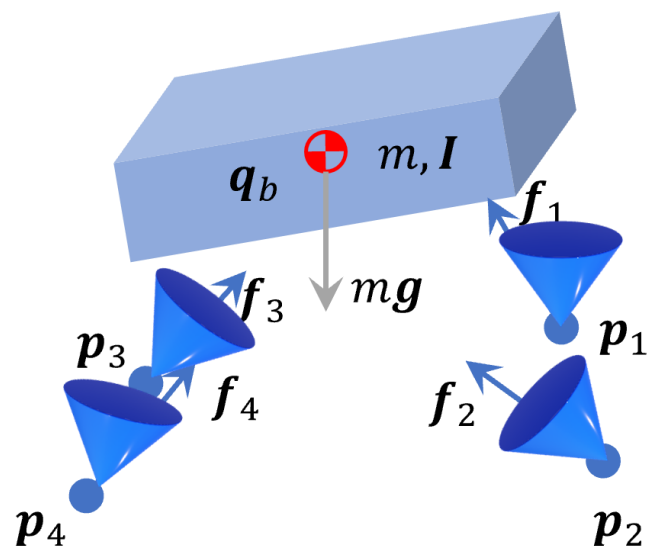
动力学模型

$p_i \in \mathcal{K}(q_b)$

几何约束

支撑相	$\mathcal{C}_i = 1$	摆动相	$\mathcal{C}_i = 0$	
	$p_i^z = \mathcal{G}(p_i^{x,y})$		$p_i^z > \mathcal{G}(p_i^{x,y})$	足端高度
	$\dot{p}_i = 0$			足端速度
	$f_i \cdot n(p_i^{x,y}) > 0$			
	$f_i \in \mathcal{F}(\mu, n)$		$f_i = 0$	足端力
	$f_i \in \mathcal{A}(q_b, p_i)$			

地形对运动的直接影响体现在足端位置与支撑力



单刚体动力学模型

$$\ddot{q}_{b_L} = \frac{1}{m} \sum_{i=1}^N f_i + g$$

$$\ddot{q}_{b_R} = I^{-1} \sum_{i=1}^N (q_{b_L} - p_i) \times f_i$$

优化变量

$$q_b(t) \in \mathbb{R}^6, p_i(t) \in \mathbb{R}^3, f_i(t) \in \mathbb{R}^3, c_i(t) \in \{0,1\}$$

约束条件

$$\ddot{q}_b = \mathcal{D}(q_b, p_{i=1,\dots,N}, f_{i=1,\dots,N}) \quad \text{动力学模型}$$

$$p_i \in \mathcal{K}(q_b) \quad \text{几何约束}$$

支撑相 $c_i = 1$

摆动相 $c_i = 0$

$$p_i^z = g(p_i^{x,y})$$

~~$$p_i^z > g(p_i^{x,y})$$~~

足端高度

$$\dot{p}_i = 0$$

足端速度

~~$$f_i \cdot n(p_i^{x,y}) > 0$$~~

~~$$f_i \in \mathcal{T}(p_i, n)$$~~

$$f_i = 0$$

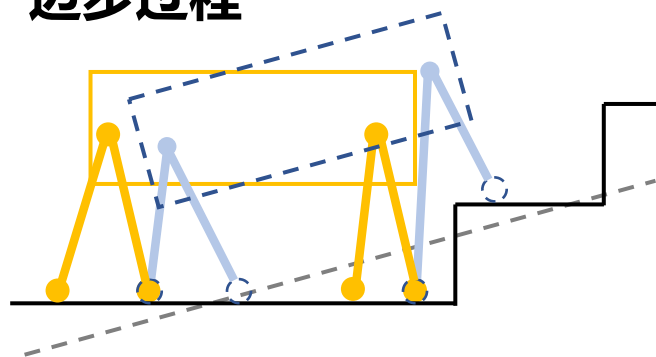
足端力

~~$$f_i \in \mathcal{A}(q_b, p_i)$$~~

在大部分**低速**且**落脚点合理**的工况下，地形主要影响脚的**支撑位置（落脚点）**

当**落地点固定**时，**地形消失**。复杂地形运动可以**拆分**简化为若干落脚点转移过程（**迈步过程**）

迈步过程



优化变量

$$\mathbf{q}_b^k(t) \in \mathbb{R}^6, \mathbf{f}_i^k(t) \in \mathbb{R}^3, t \in [t^k, t^{k+1}]$$

约束条件

$$\ddot{\mathbf{q}}_b^k = \mathcal{D}(\mathbf{q}_b^k, \mathbf{p}_{i=1,\dots,N}^k, \mathbf{f}_{i=1,\dots,N}^k)$$

动力学模型

$$\{\mathbf{q}_b, \dot{\mathbf{q}}_b, \mathbf{f}_i\}^k(t^{k+1}) = \{\mathbf{q}_b, \dot{\mathbf{q}}_b, \mathbf{f}_i\}^{k+1}(t^{k+1})$$

始末状态

单刚体动力学模型

$$\ddot{\mathbf{q}}_{b_L} = \frac{1}{m} \sum_{i=1}^N \mathbf{f}_i + \mathbf{g}$$

$$\ddot{\mathbf{q}}_{b_R} = \mathbf{I}^{-1} \sum_{i=1}^N (\mathbf{q}_{b_L} - \mathbf{p}_i) \times \mathbf{f}_i$$

支撑相 $\mathcal{C}_i = 1$

摆动相 $\mathcal{C}_i = 0$

$$\mathbf{f}_i \cdot \mathbf{n}(\mathbf{p}_i^{x,y}) > 0$$

$$\mathbf{f}_i = 0$$

足端力

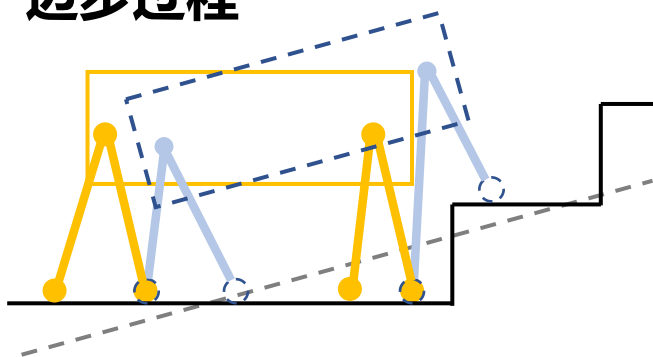
$$\mathbf{f}_i \in \mathcal{F}(\mu, n)$$

$$\mathbf{f}_i \in \mathcal{A}(\mathbf{q}_b, \mathbf{p}_i)$$

在分析**泛化**能力时，对不同**地形**的适应转化为对不同**迈步过程**的适应

复杂地形运动可以**拆分**简化为若干落脚点转移过程（**迈步过程**）

迈步过程



优化变量

$$\mathbf{q}_b^k(t) \in \mathbb{R}^6, \mathbf{f}_i^k(t) \in \mathbb{R}^3, t \in [t^k, t^{k+1}]$$

约束条件

$$\ddot{\mathbf{q}}_b^k = \mathcal{D}(\mathbf{q}_b^k, \mathbf{p}_{i=1,\dots,N}^k, \mathbf{f}_{i=1,\dots,N}^k)$$

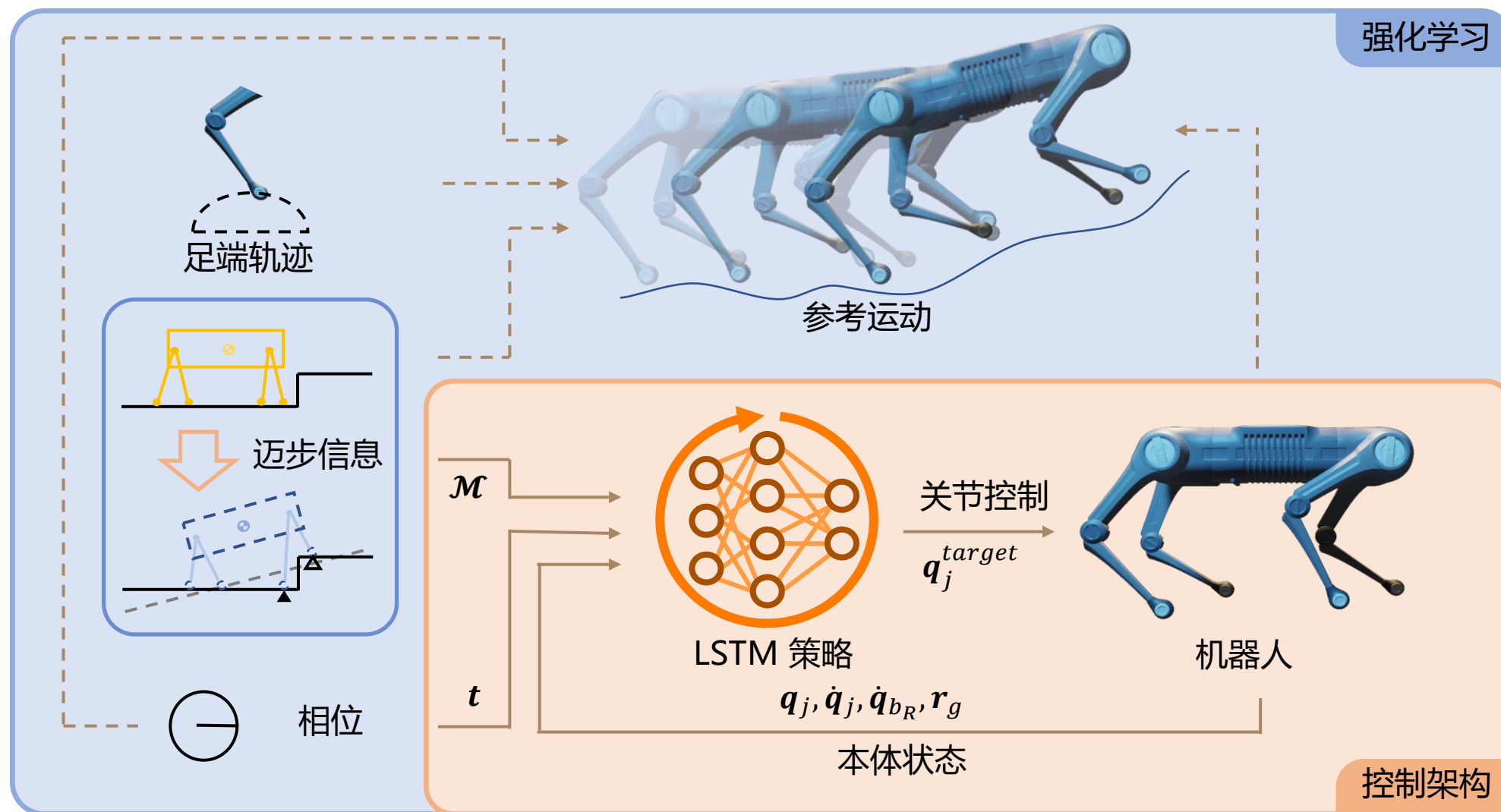
动力学模型

$$\{\mathbf{q}_b, \dot{\mathbf{q}}_b, \mathbf{f}_i\}^k(t^{k+1}) = \{\mathbf{q}_b, \dot{\mathbf{q}}_b, \mathbf{f}_i\}^{k+1}(t^{k+1})$$

始末状态

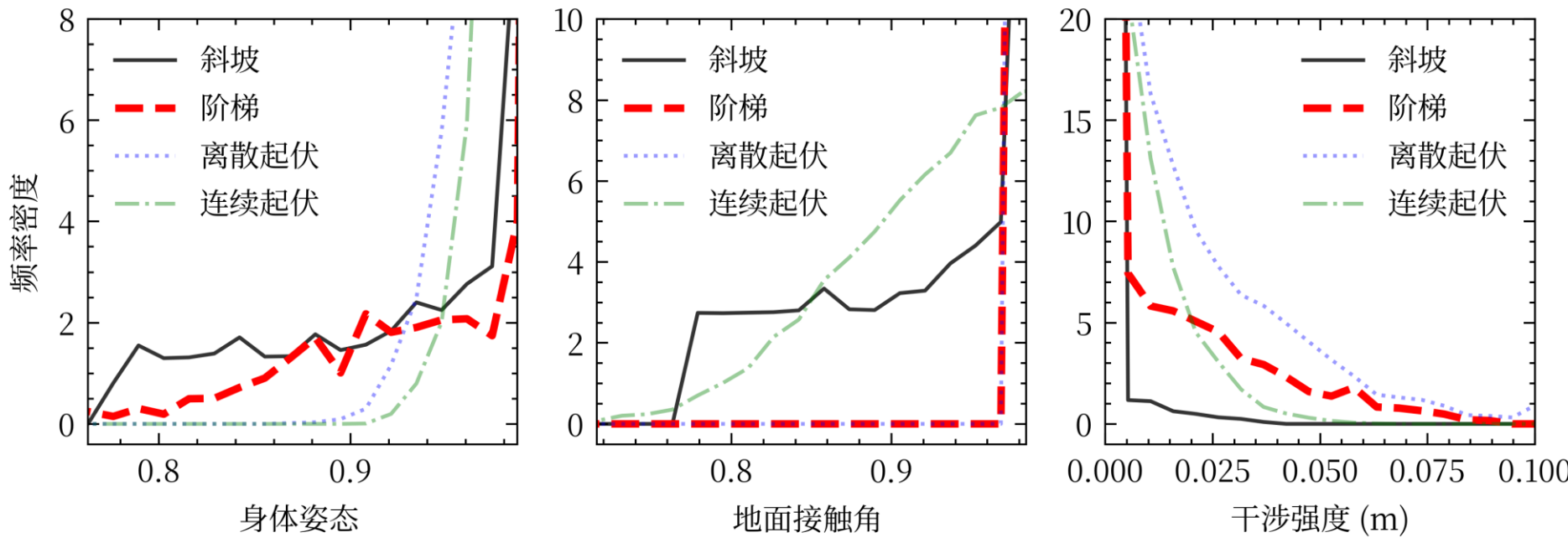
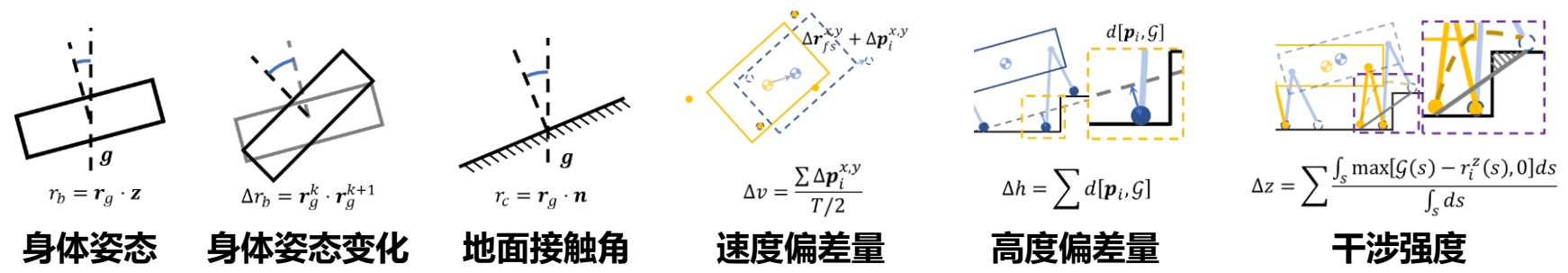
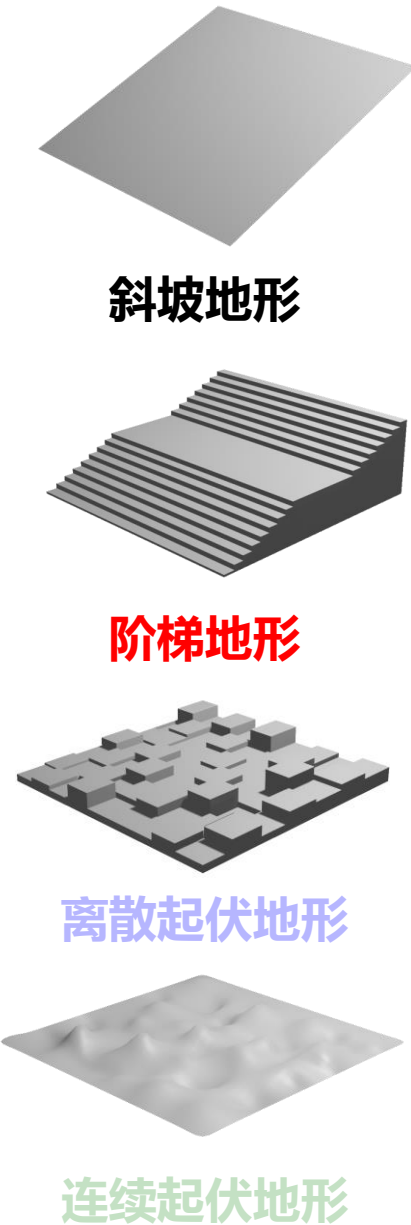


从**运动学**的角度来看，**任意地形**都可以近似为**梅花桩**地形



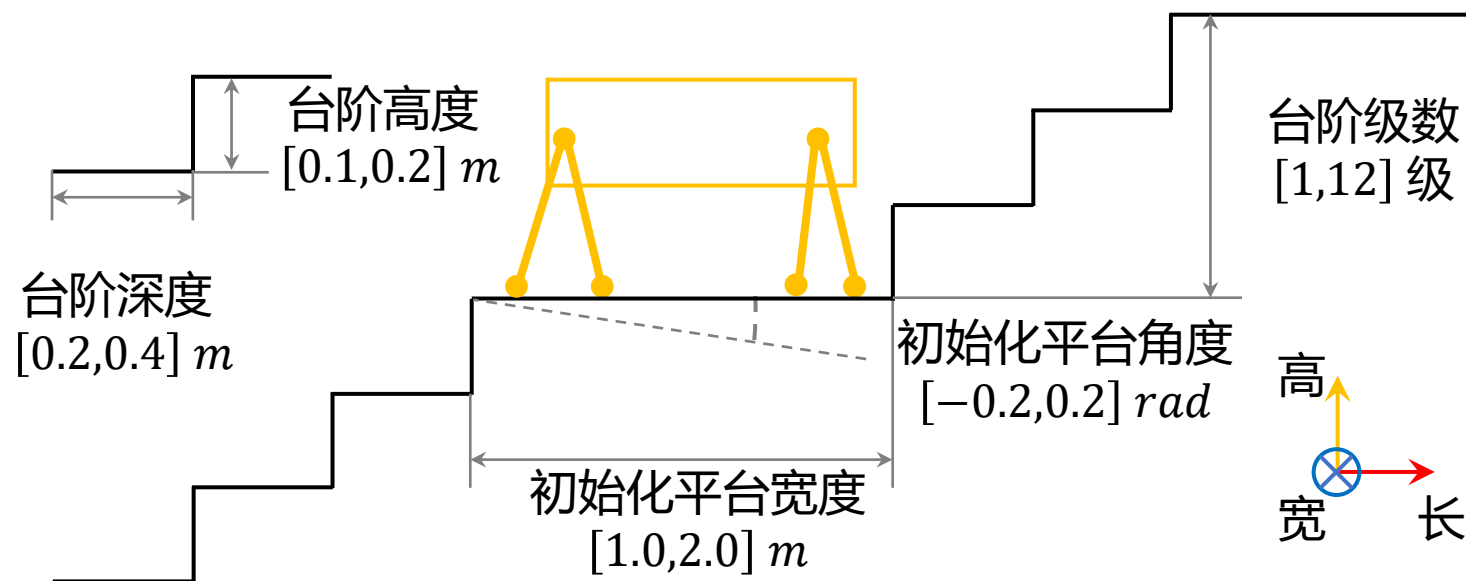
使用**迈步**过程的**始末状态**代替地形，作为**引导**信号，并生成参考运动

迈步过程工况指标

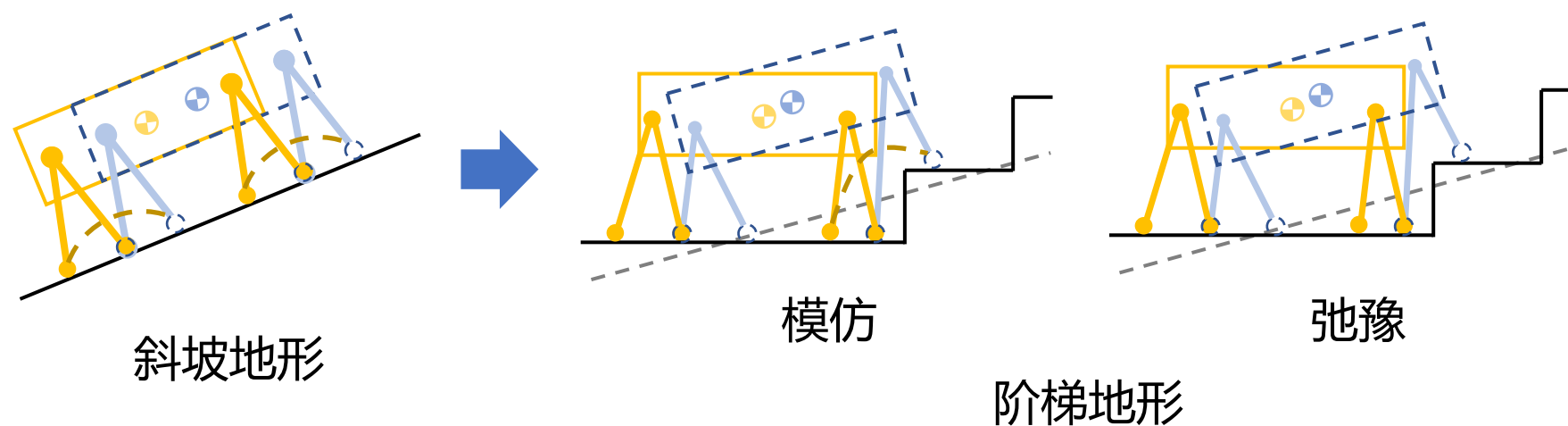


阶梯地形和斜坡地形的组合训练样本覆盖的工况相对较为广泛

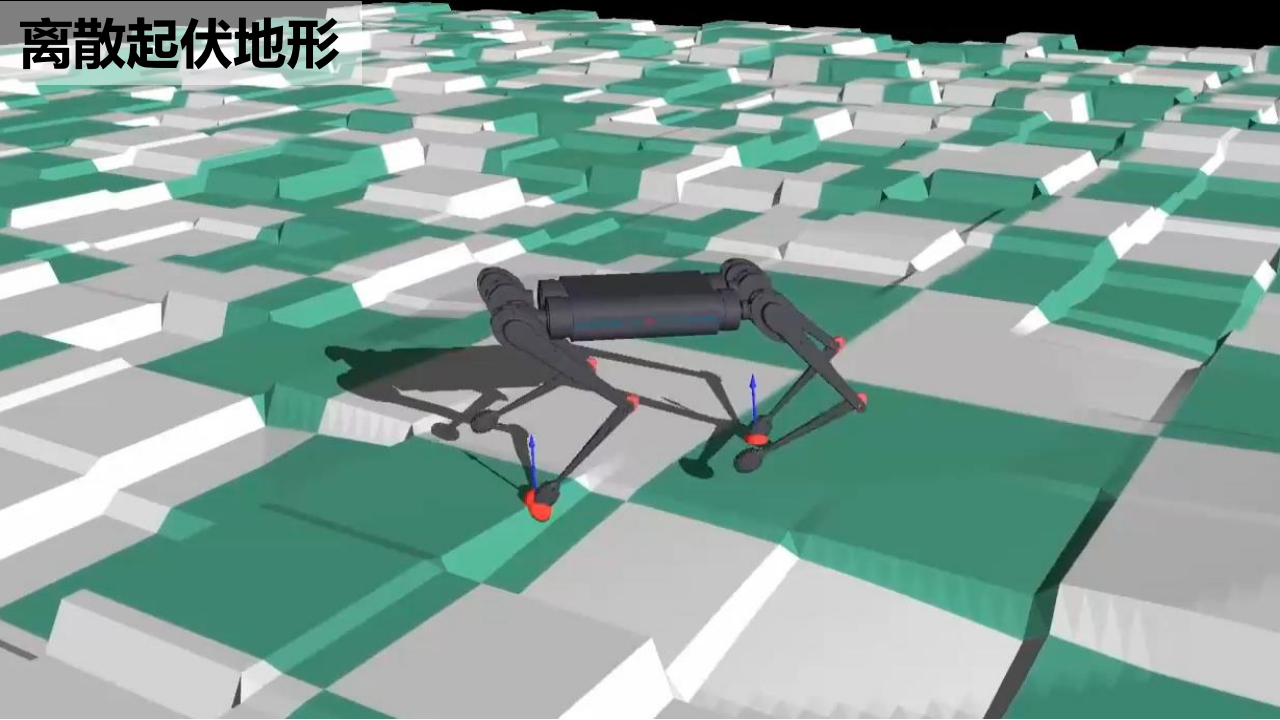
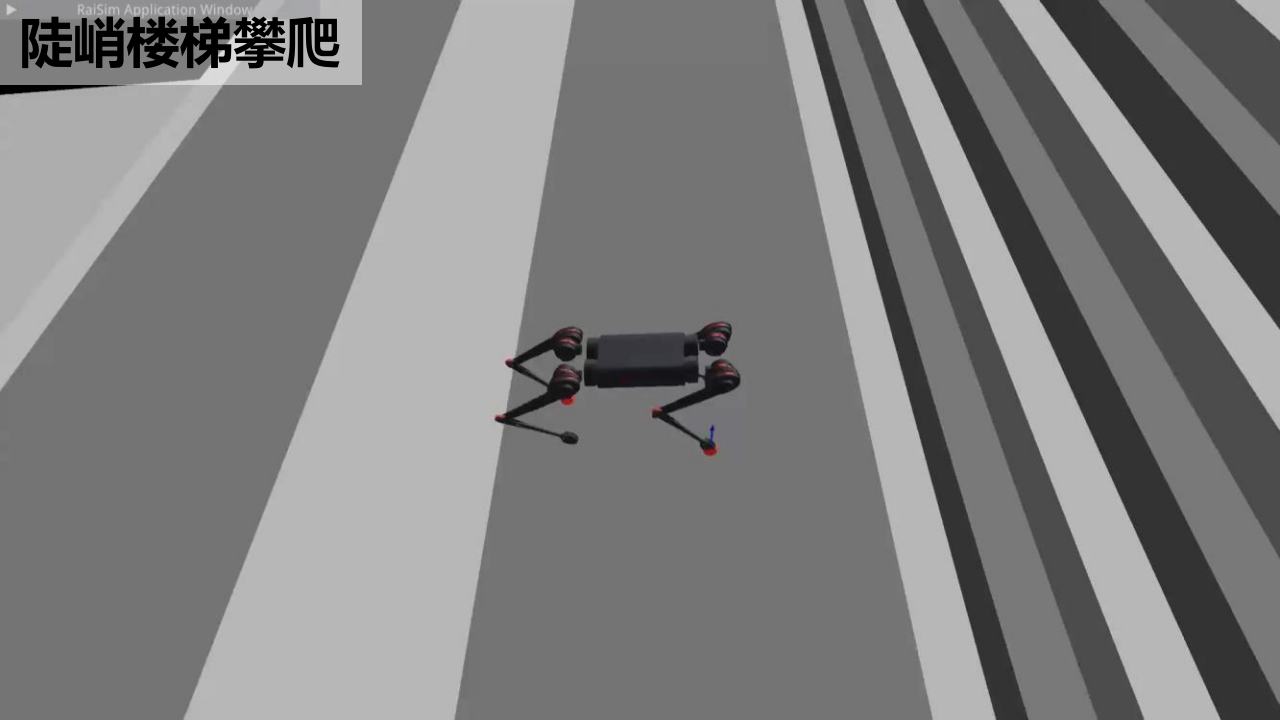
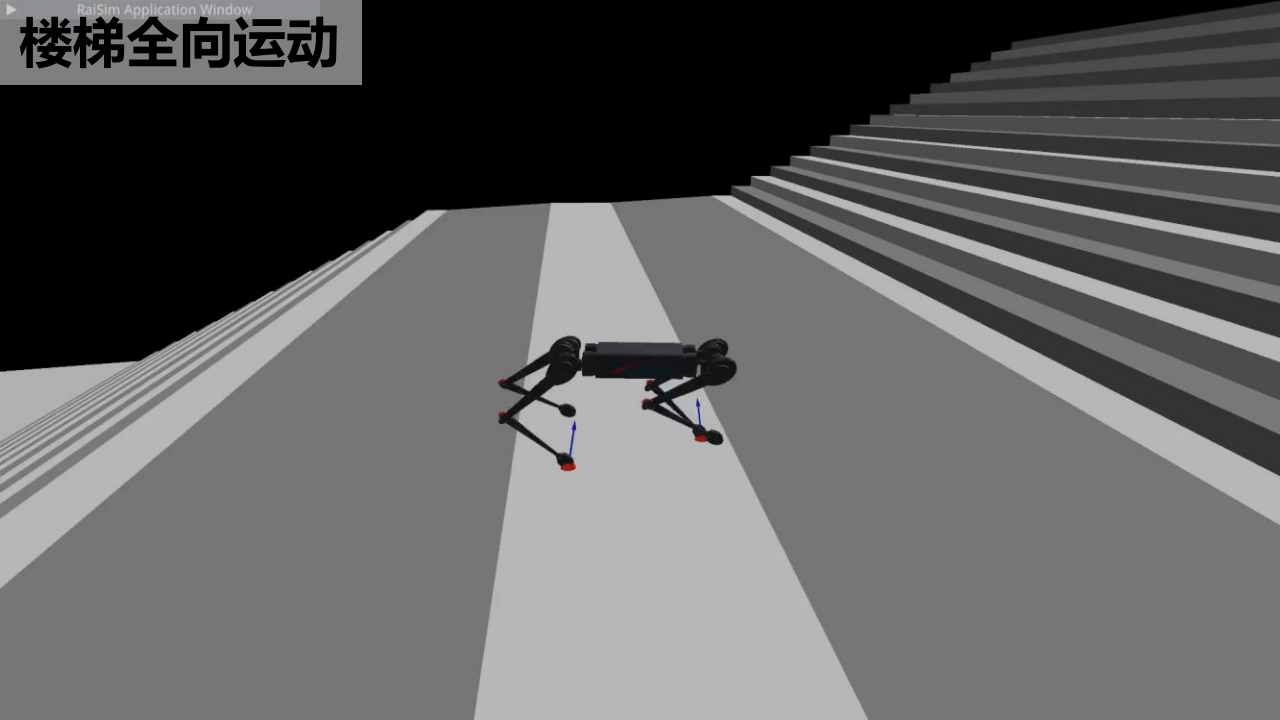
训练地形



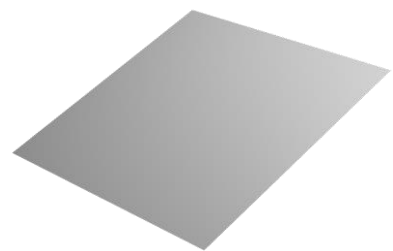
三阶段训练



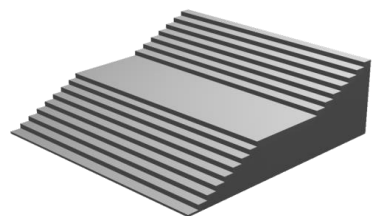
RaiSim Application Window
楼梯全向运动



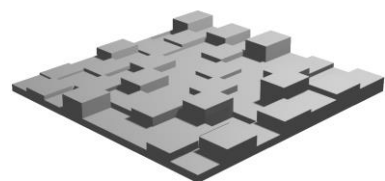
不同地形训练的控制器在地形适应能力



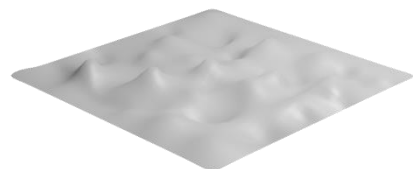
斜坡地形



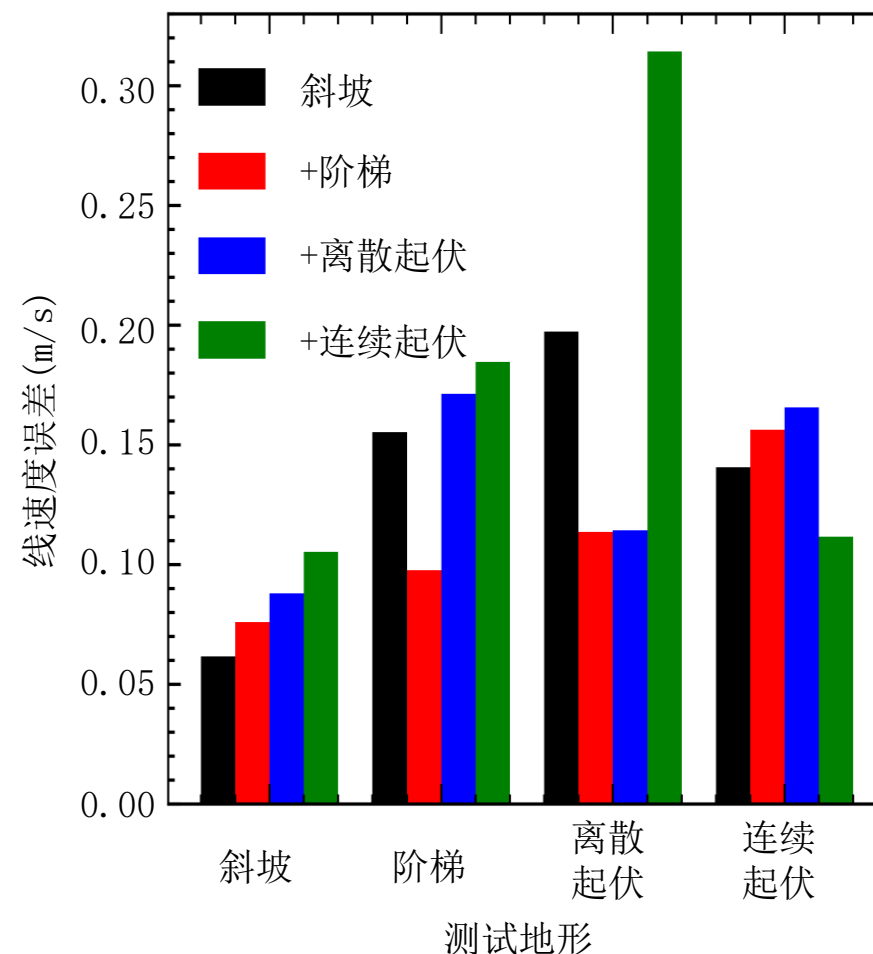
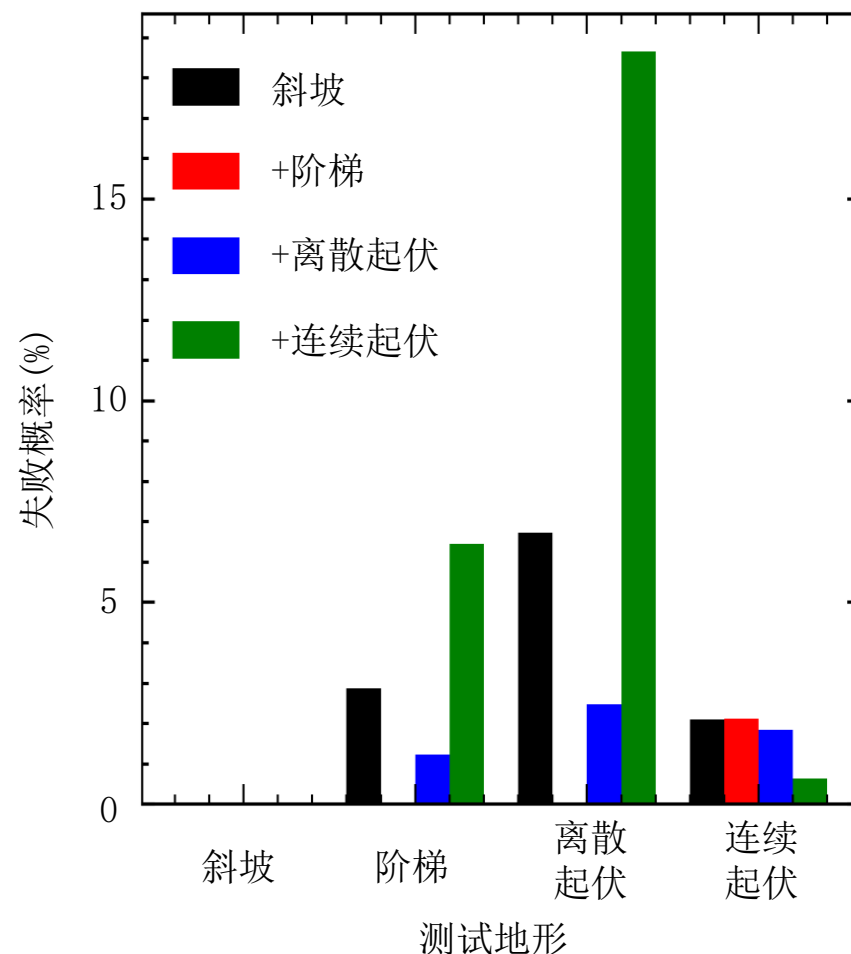
阶梯地形



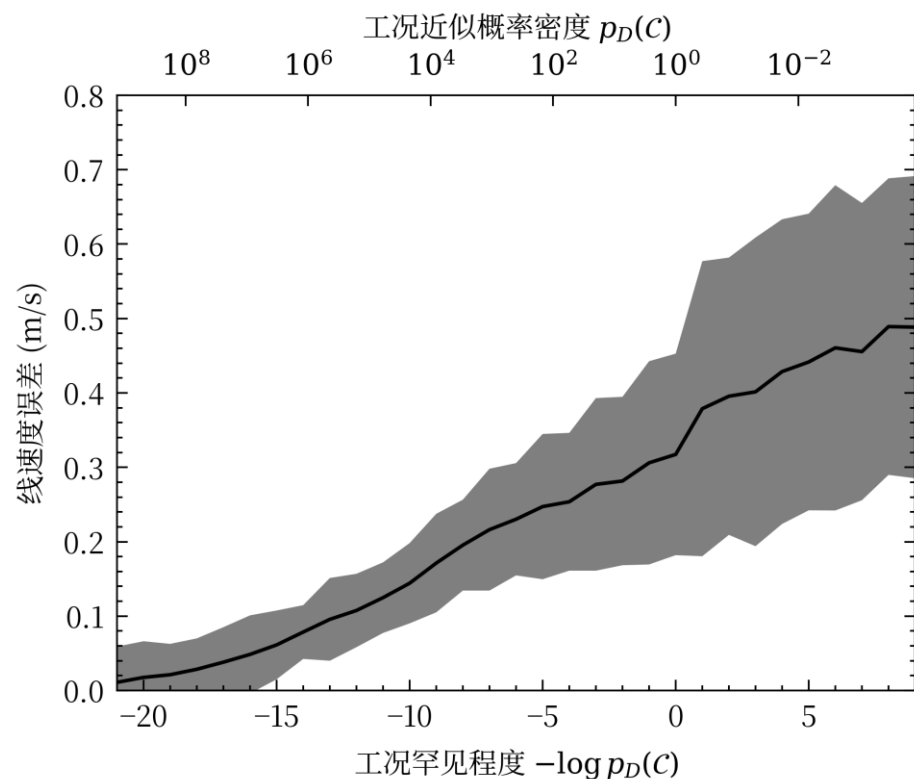
离散起伏地形



连续起伏地形

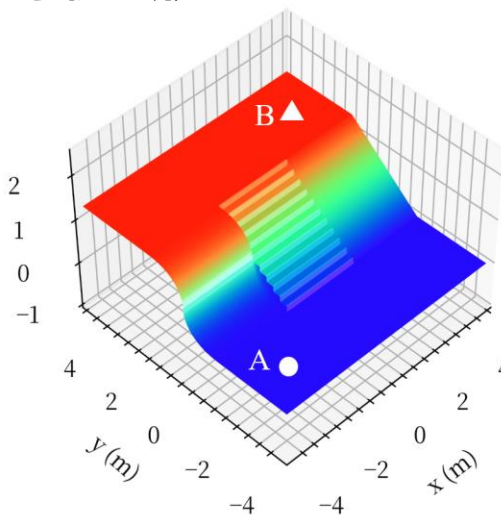


基于**阶梯地形**训练的控制器在地形适应能力**强于**其他地形训练得到的控制器

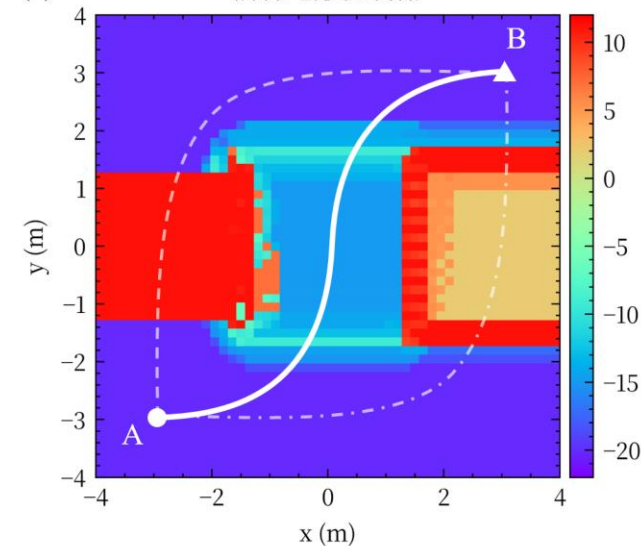


- 工况**出现频率**与**速度误差**负相关
- 工况**罕见程度**可以预测控制器**通过性**

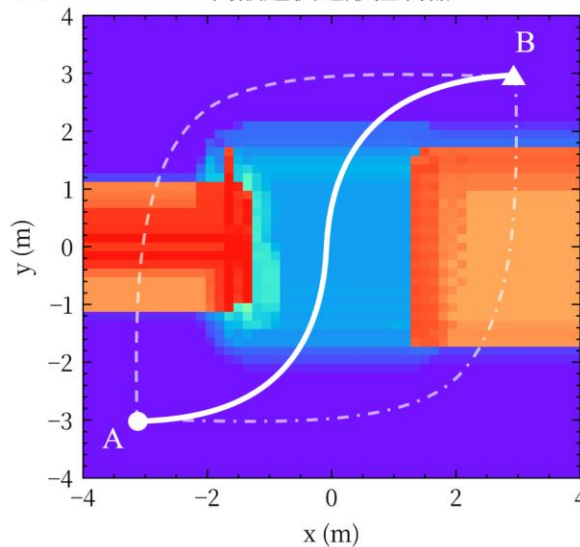
(a) 示例地形



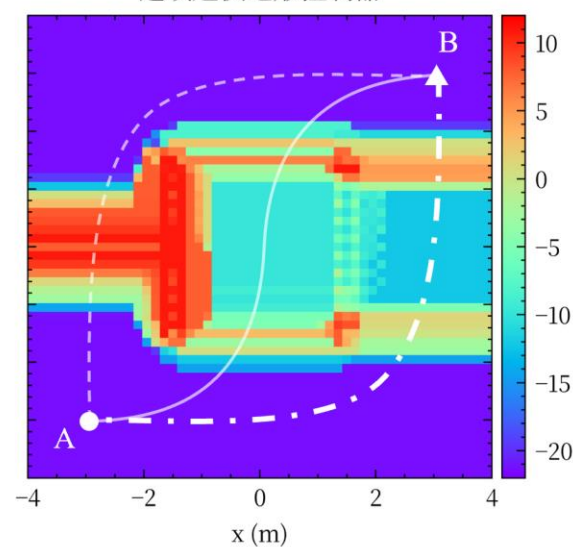
(b) 阶梯地形控制器



(c) 离散起伏地形控制器



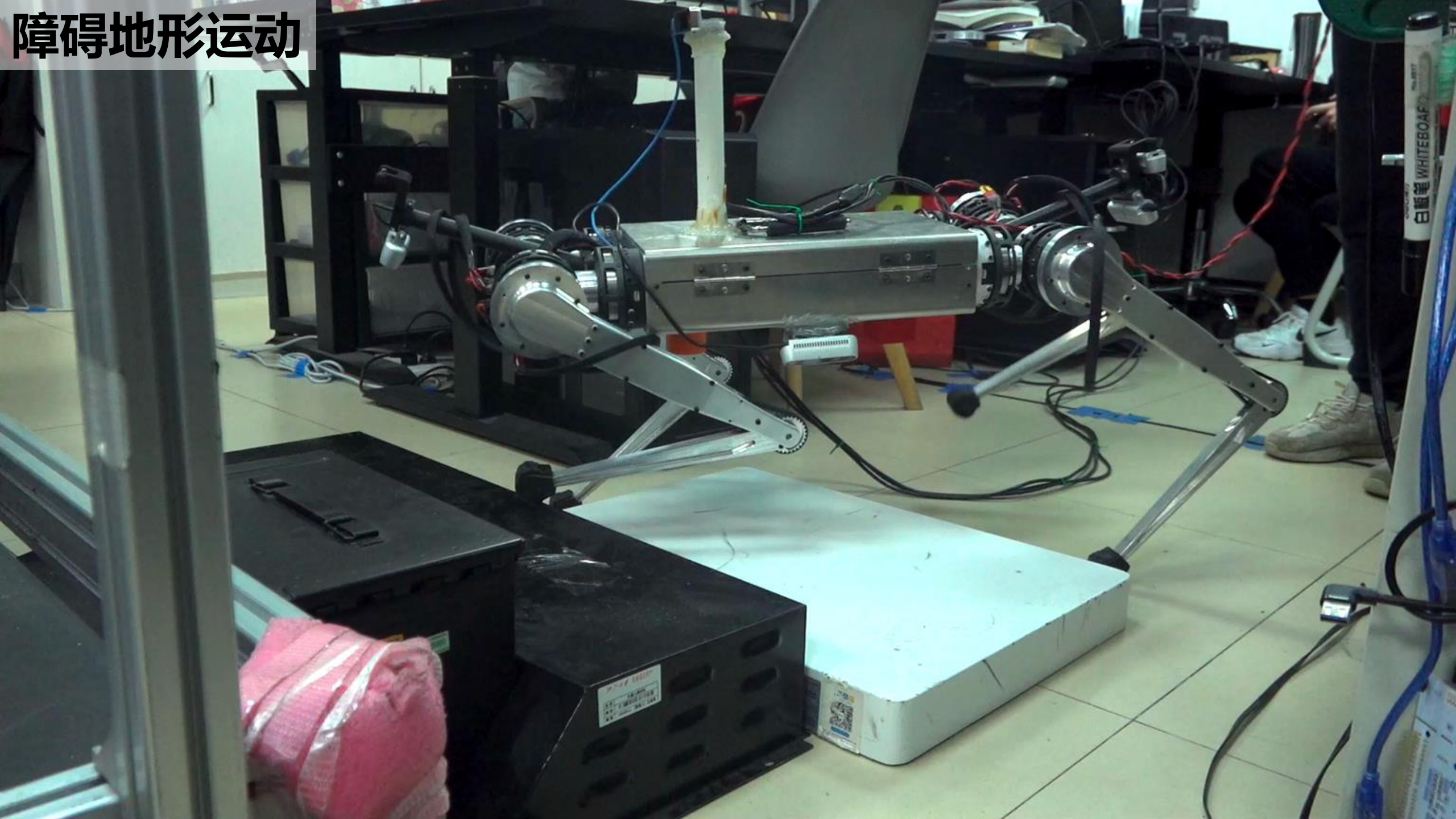
(d) 连续起伏地形控制器



楼梯地形运动

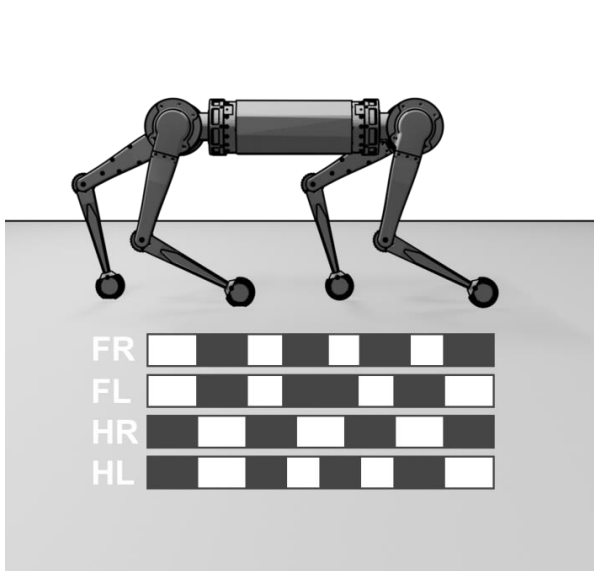


障碍地形运动

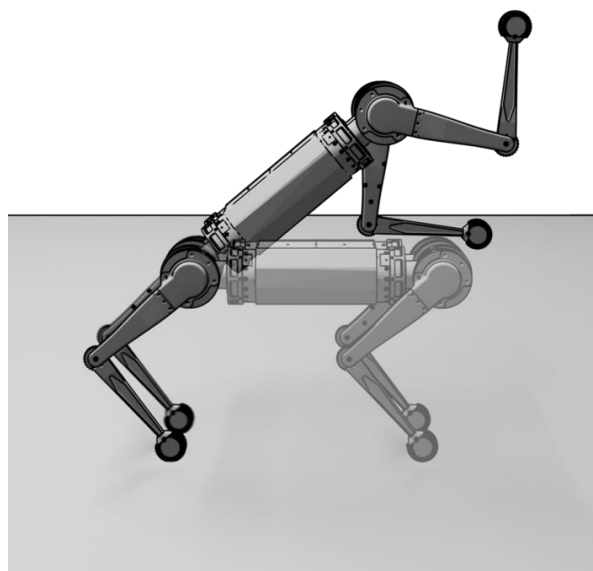


针对足式运动强化学习控制中的**多任务学习**问题与**泛化迁移**问题，展开了三项具体研究：

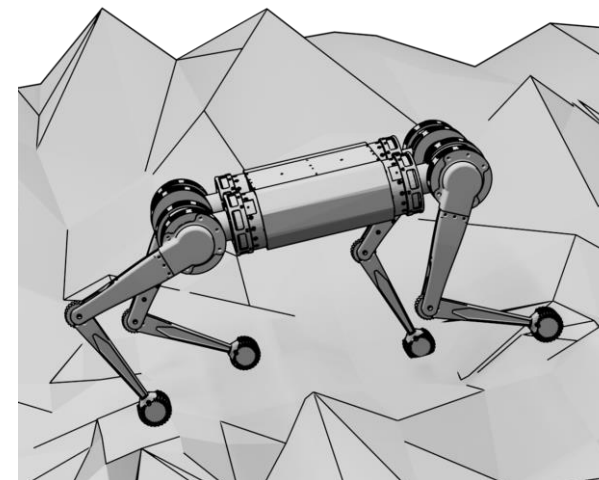
1. 发展了基于相位引导的**步态切换**控制方法，揭示了**引导信号**在多任务学习中的作用；
2. 发展了针对**一般动作**的模仿学习控制方法，揭示了**未来运动信息**对控制的影响机理；
3. 发展了基于落脚点引导的**地形适应**控制方法，提出了基于工况的**泛化能力分析**方法。



步态切换



复杂动作



地形适应

1. **人机协同**：建立人机共融系统，使用人的**实时运动**数据控制机器人，将人类**智能**与机器人**机能**融合，推动机器人的功能发展；
2. **理论分析**：参考轨迹**可预测性**理论只具有**信息学**含义，不具备动力学含义，未来从**动力学**的角度解释不同运动轨迹的**差别**；
3. **自主控制**：目前的**环境适应**方法仅限于对地形几何形貌的适应，未来将以语言模型为基础，构建**多模态模型**，挖掘机器人**理解场景**的能力。

**感谢答辩委员会专家
请批评指正**