上海大学未来技术学院 | 上海大学人工智能研究院
SCHOOL OF FUTURE TECHNOLOGY, SHANGHAI UNIVERSITY | INSTITUTE OF ARTIFICIAL INTELLIGENCE, SHANGHAI UNIVERSITY

THU-SHU ROBOART LAB

足式机器人强化学习控制

叶林奇

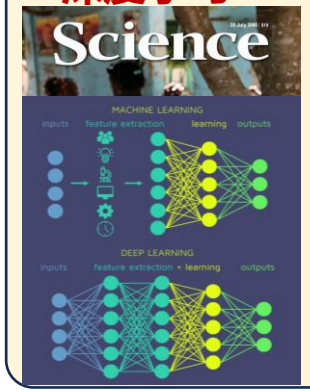
2024.4.10

提纲

THU-SHU ROBOART LAB

人工智能取得巨大成功，无人系统的应用初见成效

Hinton提出
深度学习



2006

游戏
AI

AlphaGo战胜
人类最强围棋选手

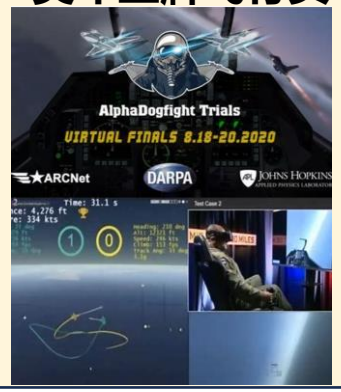
2016

四足机器人强化
学习超越传统控制

2019

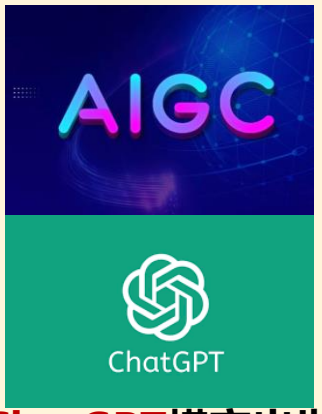
AI星际争霸超越
99.8%人类玩家

模拟空战AI打败
美军王牌飞行员



2020

大
模型



ChatGPT横空出世，
AIGC火爆全球

2022

无人机竞速AI战胜
人类世界冠军

2023

谷歌发布具身多模态
大语言模型PaLM-E

智能
无人系统

Why Learning?

学习方法在多个无人系统领域引起巨大变革，不断突破传统方法上限

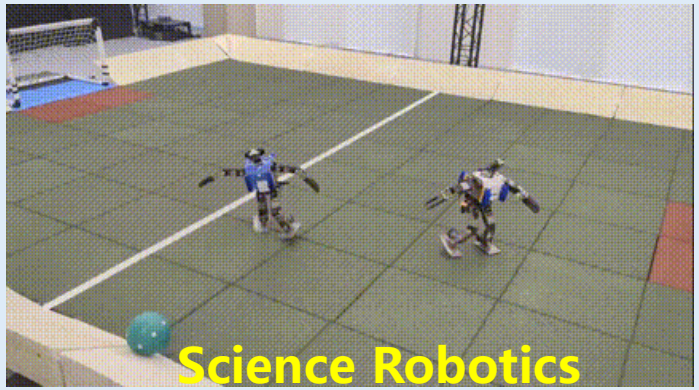
2021 → 2022 → 2023

四足机器人

- 4分钟学会走路，可迁移泛化，适应野外各种地形
- 获DARPA机器人地下挑战赛冠军，全程无一次摔倒

双足机器人

- 运球、射门、防守、跌倒爬起
- 1v1实物验证，10天完成训练
- 3v3仿真验证，50天完成训练
- 分阶段学习，教师+自博弈



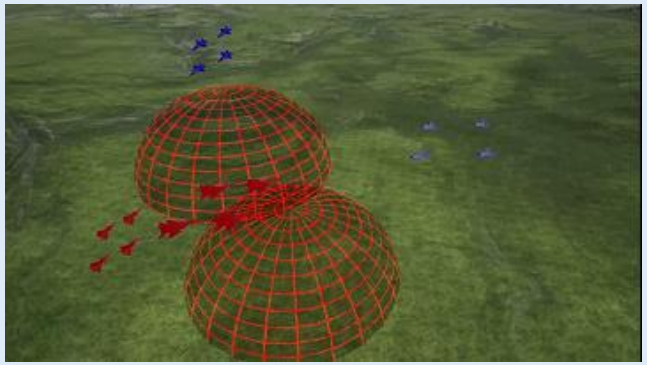
无人机

- 学习50分钟，时速100千米
- 战胜人类冠军并创造最快比赛纪录
- 机器智能领域的一个里程碑

学习方法能够应对近乎无穷的复杂度，快速超越并具备持续进化能力

学习方法应用于真实物理系统成功的关键在于虚实融合

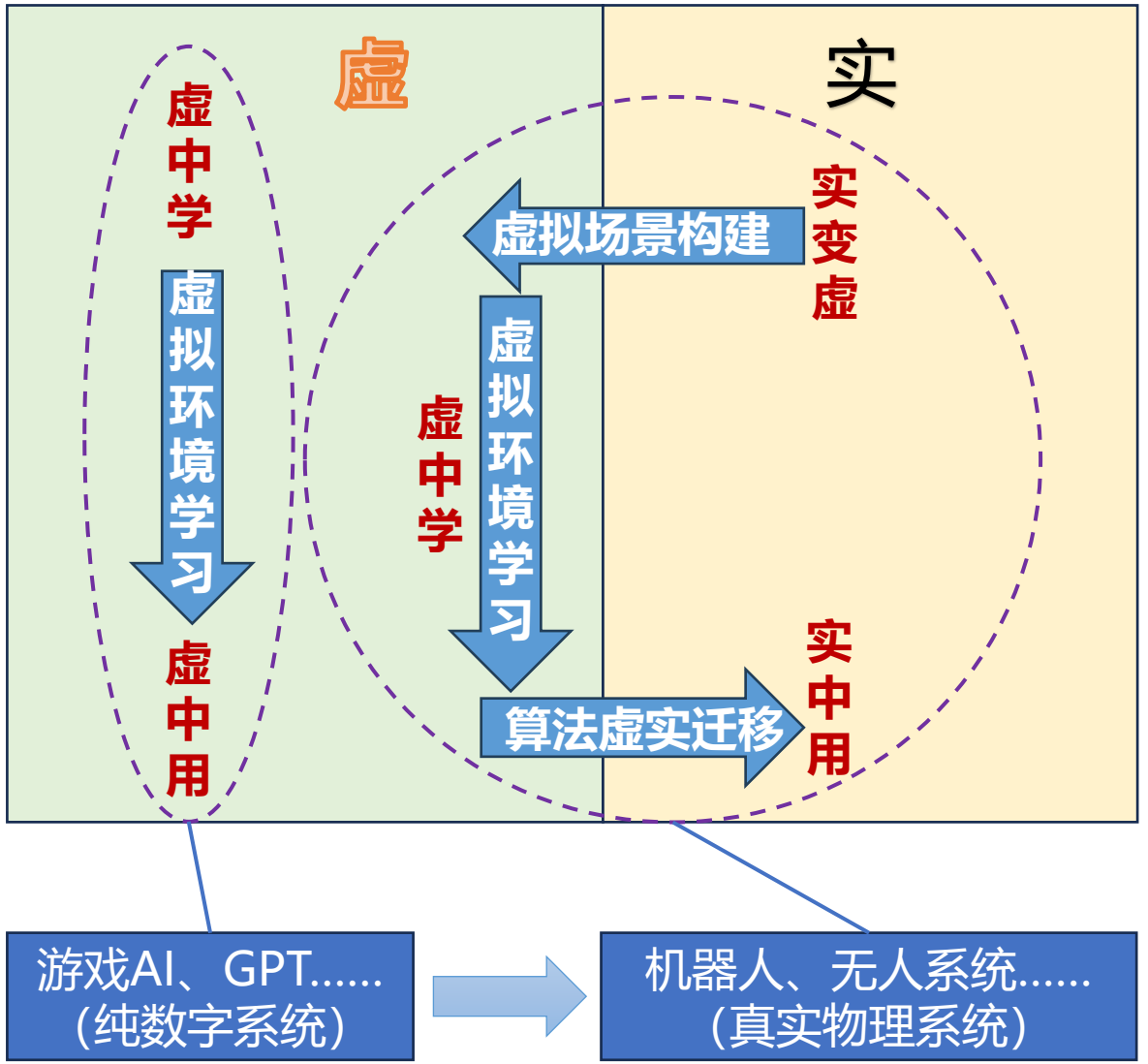
虚拟环境学习



- **优点：**风险小、成本低、效率高
- **缺点：**实物的有效性难以保证

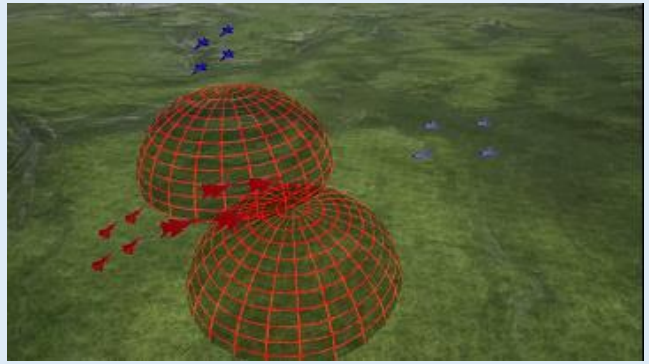
实物在线学习

- **优点：**真实有效
- **缺点：**风险大、成本高、周期长



虚实融合学习三要素：精确建模、高效学习、有效迁移

虚拟环境学习



- **优点：**风险小、成本低、效率高
- **缺点：**实物的有效性难以保证

实物在线学习

- **优点：**真实有效
- **缺点：**风险大、成本高、周期长

真实物理系统 虚实融合学习

- **高保真建模(实→虚)**
  - 基于监督学习的执行器模型
  - 基于真实数据的噪声模型……
- **高效学习训练(虚+实)**
  - 教师-学生学习架构
  - 分层学习……
- **算法有效迁移(虚→实)**
  - 域随机化
  - 迁移学习……

具身智能：具备了身体的人工智能（机器人+小脑+大脑）

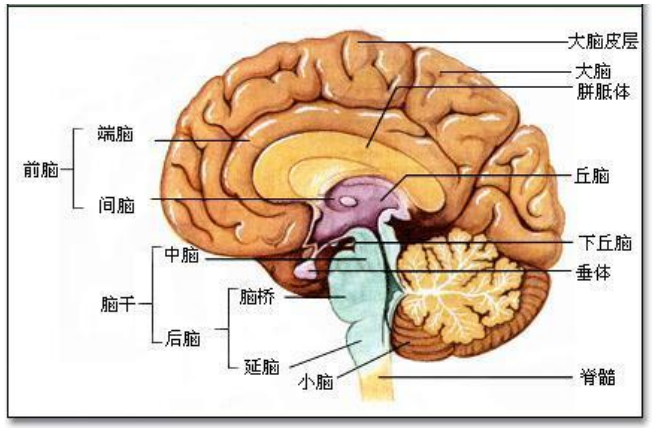
机器人+强化学习→敏捷运动小脑

3月12日，逐际动力双足机器人P1第一次来到位于深圳的郊野公园塘朗山，基于强化学习，零样本、无保护、全开放进行测试，开箱即跑，在完全陌生的野外环境高动态完成了在多种复杂地形上移动，表现出强化学习训练后，优异的控制力和稳定性。

机器人+大模型→智能决策大脑

3月14日，美国Figure机器人公司发布了第一个OpenAI大模型加持的机器人demo，视频展示了端到端神经网络框架下机器人与人类的对话，没有任何远程操作。机器人可以：描述其视觉体验；规划未来的行动；反思自己的记忆；口头解释推理过程。

人的大脑和小脑协同合作，分工配合

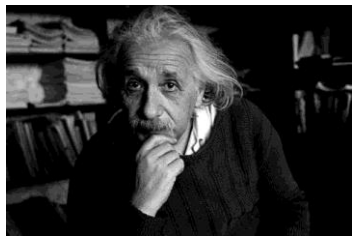


大脑（认知，知道怎么游泳）

小脑（技能，会游泳）

大脑是总司令,小脑就是参谋长。大脑在发出指令时,先将信息传递给小脑,小脑分析校正,确保准确性后,再传回到大脑,由大脑做出最终的指令。

- **大脑**支配人的生命活动：语言、运动、听觉、视觉、情感表达等。它能够调节消化、呼吸、循环、泌尿、生殖、运动等中枢。
- 智力：观察力、注意力、记忆力、思维力、想象力，大脑是一切思维活动的物质基础 。
- 能力：思维记忆、学习获得、认识理解、判断推理、综合分析、语言表达、社会活动能力、意识情绪

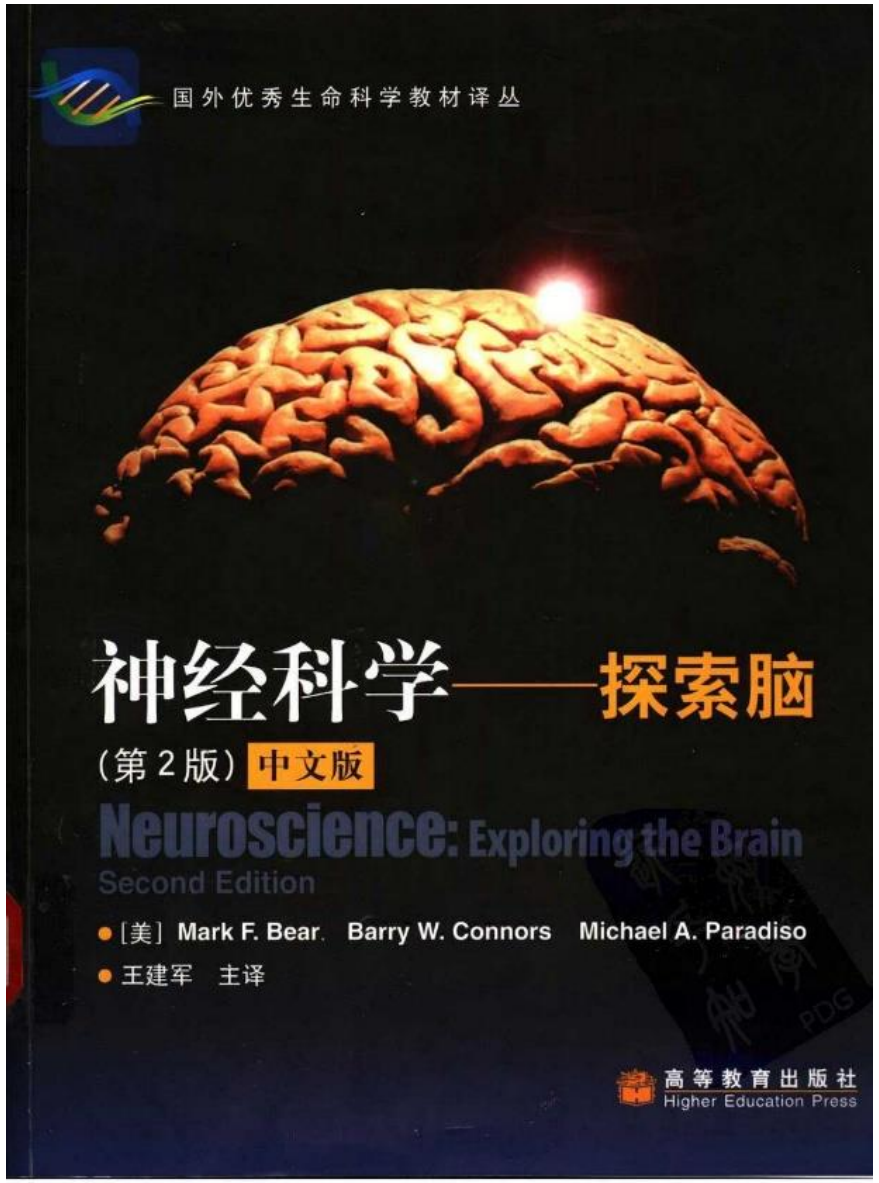


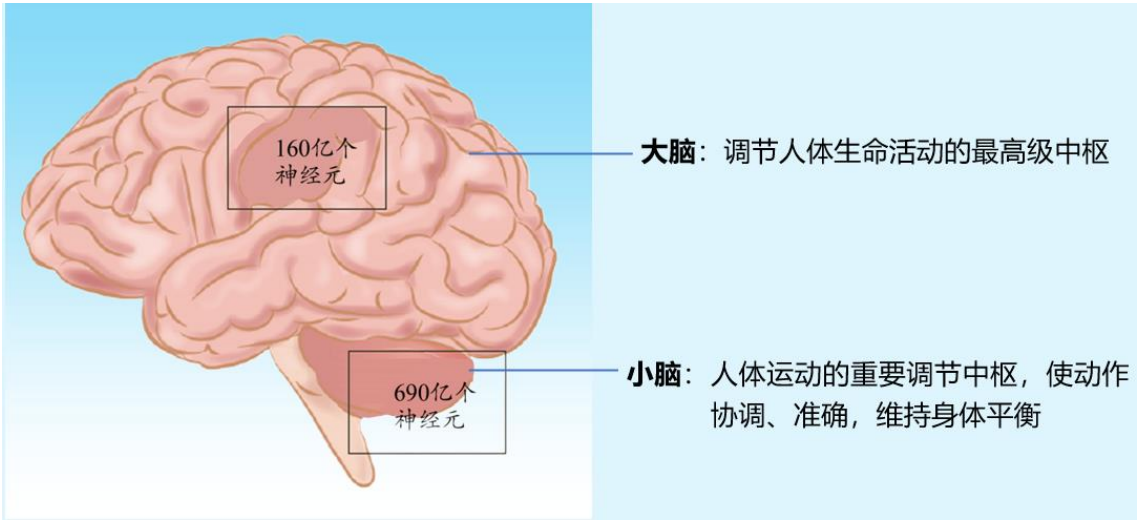


- **小脑**通过复杂的调节和反馈机制成为维持平衡和肌张力的协调中枢（**运动协调**），它还能使躯体肌肉系统完成精细的技巧性运动（**运动学习**）。小脑像计算机一样能扫描和协调感觉传入并调节运动传出。




人的小脑仅占全脑体积1/10，但神经元数量却是大脑的4倍





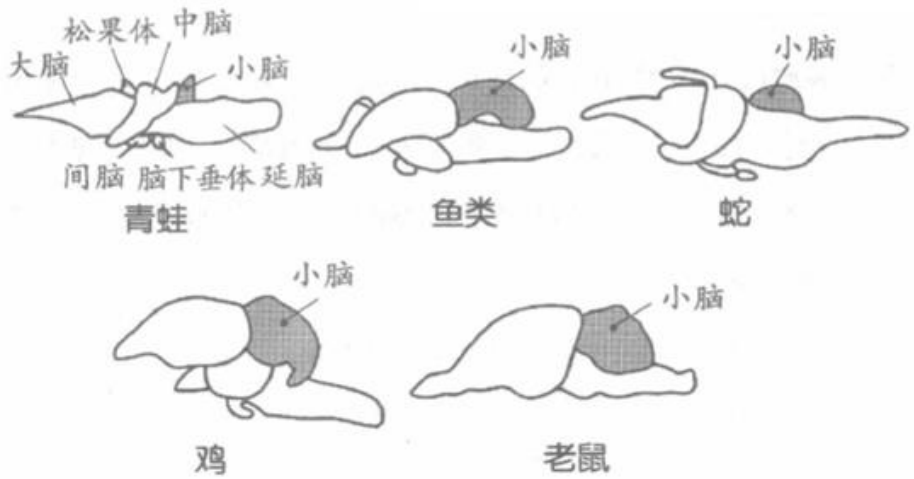
运动控制的等级

水平	功能	结构
高	运动战略	新皮层联络区、基底神经节
中	运动战术	运动皮层、小脑
低	运动执行	脑干、脊髓

动作越灵活的动物，小脑越发达

鸟类、鱼类、哺乳类的小脑占脑部的比例很大，而两栖类或爬虫类的就很小。研究表明，小脑受损患者运动控制变得粗糙、不精细、控制不足。

借鉴小脑功能机制，建立机器人运动控制小脑，将有可能使机器人运动能力实现质的飞跃。



小脑在精细运动控制、身体平衡、多关节运动协调和运动学习方面扮演至关重要的作用。

运动精确性和协调性的控制

- **平衡**：姿势控制，对外部干扰做出反应，提供前瞻性控制
- **移动**：躲避障碍和适应新环境，保持双眼凝视的稳定性
- **抓握**：握力的控制，不同目标物体的适应
- **时序**：运动协调，为肌肉活动提供正确时序
- **多关节运动控制**：对跨关节肌肉进行复合协同，调节互动力矩
- **感觉运动同步化**：参与获取序列运动的最佳内在模式，优化感觉运动参数

运动学习

- **建模**：建立运动器官和运动环境的神经内在模型并不断更新改进
- **预测**：内在模型预测产生运动指令，克服感觉反馈的延迟现像
- **自主**：让运动转变成更加自主的运动，减少运动细节对注意力的需求

提纲

THU-SHU ROBOART LAB

一、从人工智能到具身智能的转变

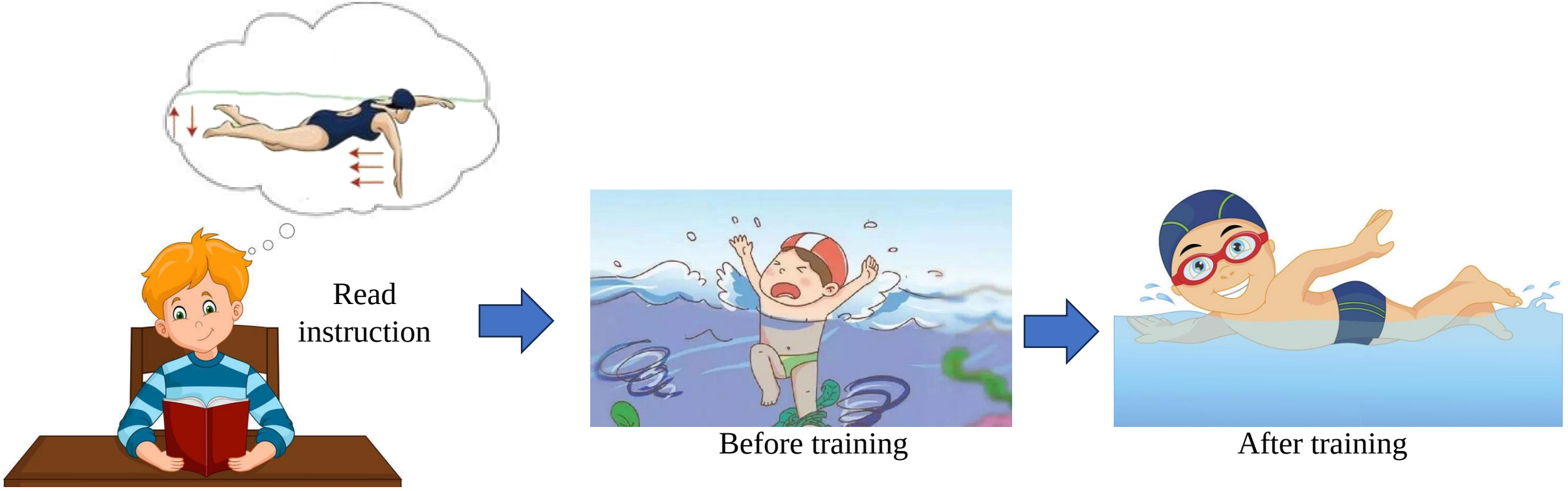
二、知识与数据双驱动的强化学习

三、双足机器人多模运动跟踪控制

四、四足机器人碰撞感知越障控制

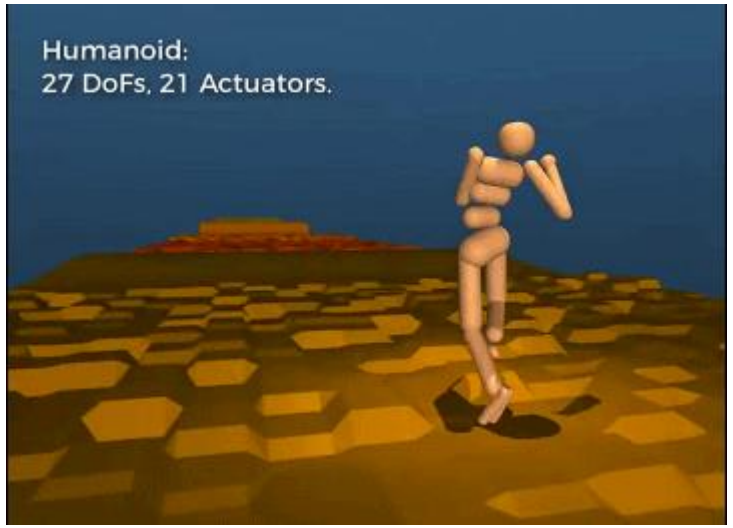
From Knowing to Doing: Learning Diverse Motion Behaviors through Instruction Learning

Linqi Ye, Jiayi Li , Yi Cheng, Xianhao Wang, Bin Liang, Yan Peng



Reinforcement Learning for Legged Locomotion

Simple reward may lead to unnatural behavior!



Deepmind 2017, Emergence of Locomotion Behaviours in Rich Environments

Rewards

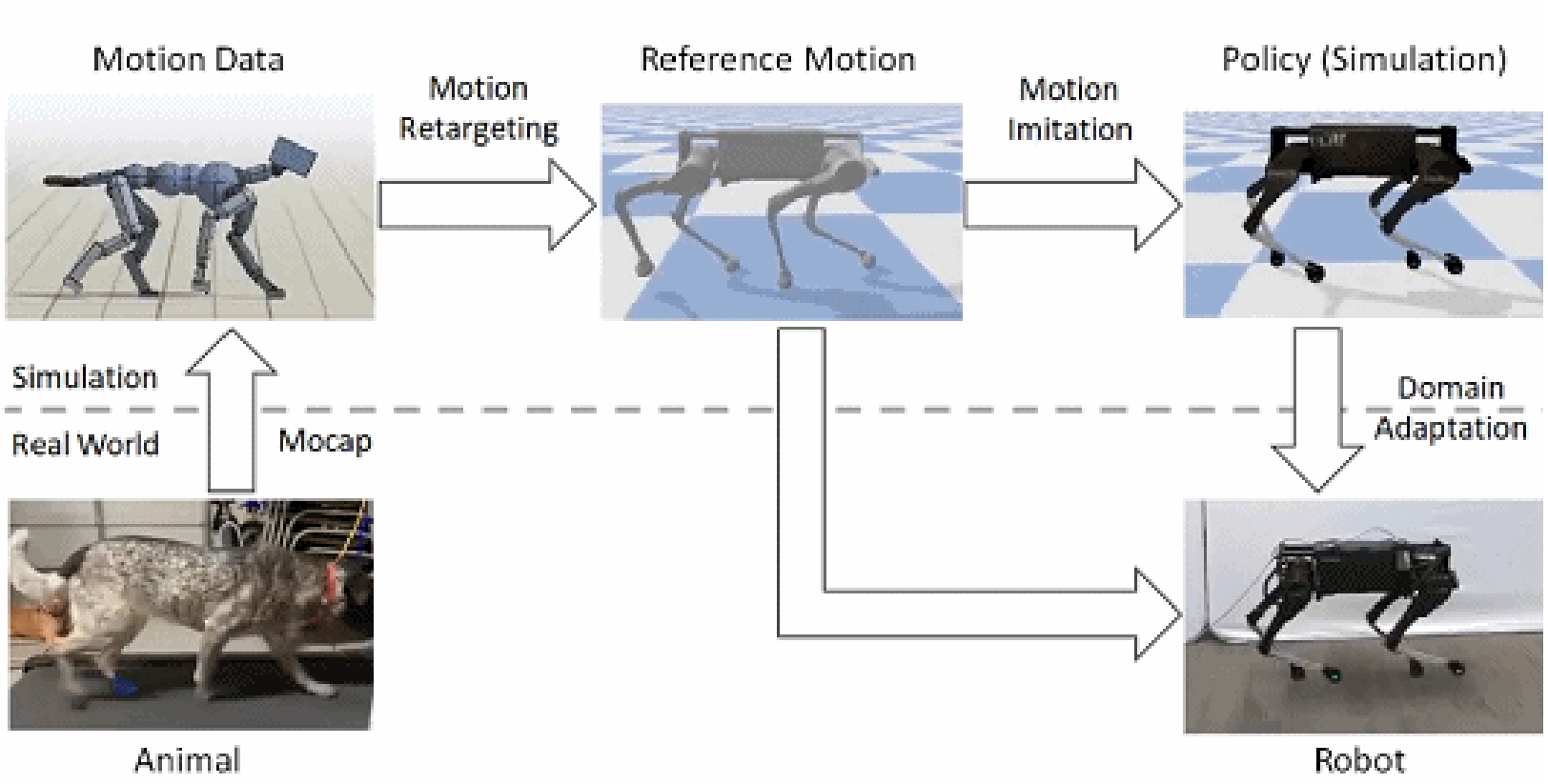
Reward for forward velocity

Energy punishment

Humanoid $r = \min(v_x, v_{\max}) - 0.005(v_x^2 + v_y^2) - 0.05y^2 - 0.02\|u\|^2 + 0.02$ where $v_{\max}$ is a cutoff for the velocity reward which we usually set to $4m/s$.

Reinforcement Learning for Legged Locomotion

Imitation learning → Follow a given reference motion



Reward Function.

$$r_t = w^{\mathrm{p}} r_t^{\mathrm{p}} + w^{\mathrm{v}} r_t^{\mathrm{v}} + w^{\mathrm{e}} r_t^{\mathrm{e}} + w^{\mathrm{rp}} r_t^{\mathrm{rp}} + w^{\mathrm{rv}} r_t^{\mathrm{rv}}$$

$$r_t^{\mathrm{p}} = \exp\left[-5 \sum_j ||\hat{\mathbf{q}}_t^j - \mathbf{q}_t^j||^2\right]$$

$$r_t^{\mathrm{v}} = \exp\left[-0.1 \sum_j ||\hat{\dot{\mathbf{q}}}_t^j - \dot{\mathbf{q}}_t^j||^2\right]$$

$$r_t^{\mathrm{e}} = \exp\left[-40 \sum_e ||\hat{\mathbf{x}}_t^e - \mathbf{x}_t^e||^2\right]$$

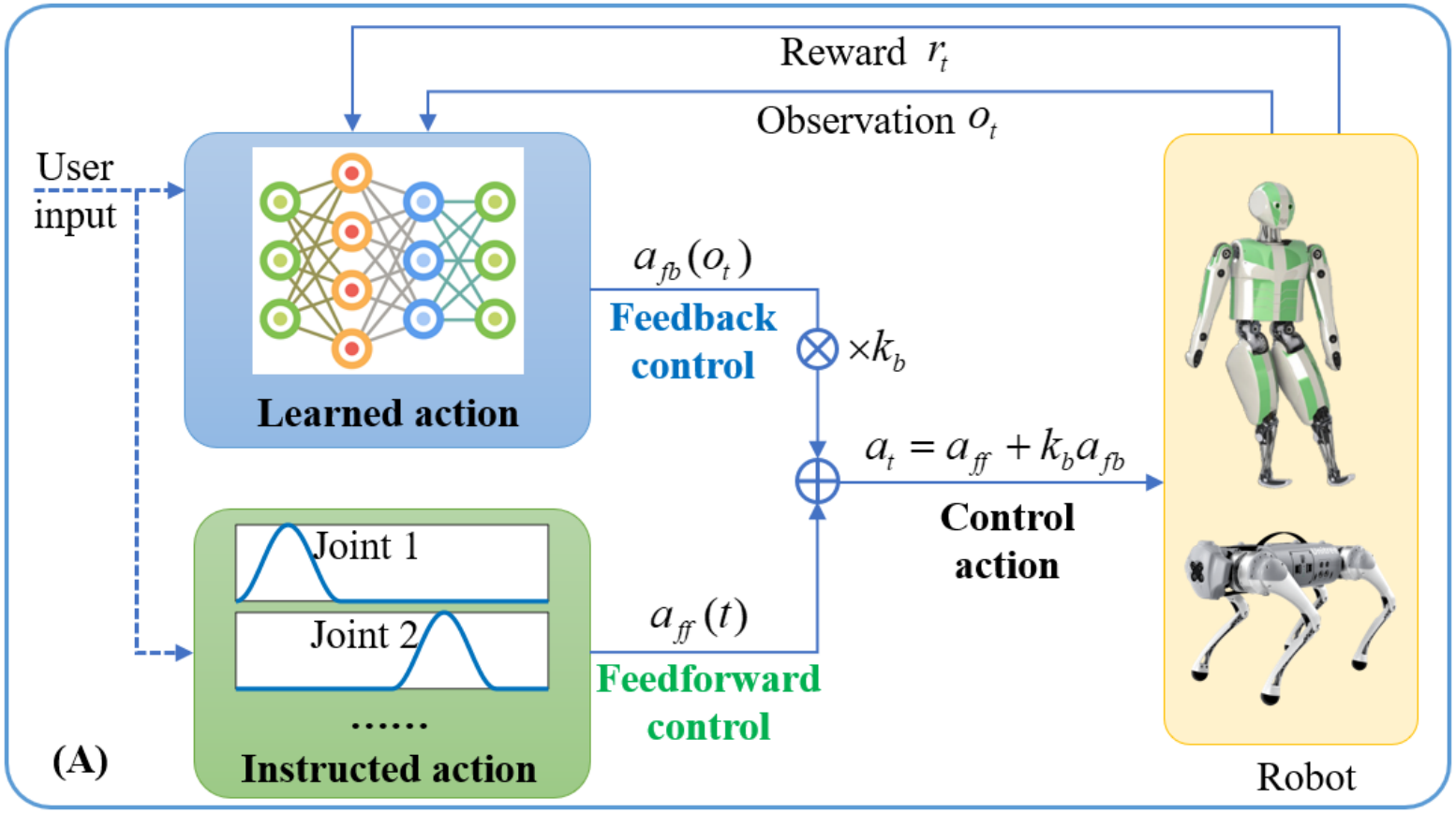
$$r_t^{\mathrm{rp}} = \exp\left[-20||\hat{\mathbf{x}}_t^{\mathrm{root}} - \mathbf{x}_t^{\mathrm{root}}||^2 - 10||\hat{\mathbf{q}}_t^{\mathrm{root}} - \mathbf{q}_t^{\mathrm{root}}||^2\right]$$

$$r_t^{\mathrm{rv}} = \exp\left[-2||\hat{\dot{\mathbf{x}}}_t^{\mathrm{root}} - \dot{\mathbf{x}}_t^{\mathrm{root}}||^2 - 0.2||\hat{\dot{\mathbf{q}}}_t^{\mathrm{root}} - \dot{\mathbf{q}}_t^{\mathrm{root}}||^2\right]$$

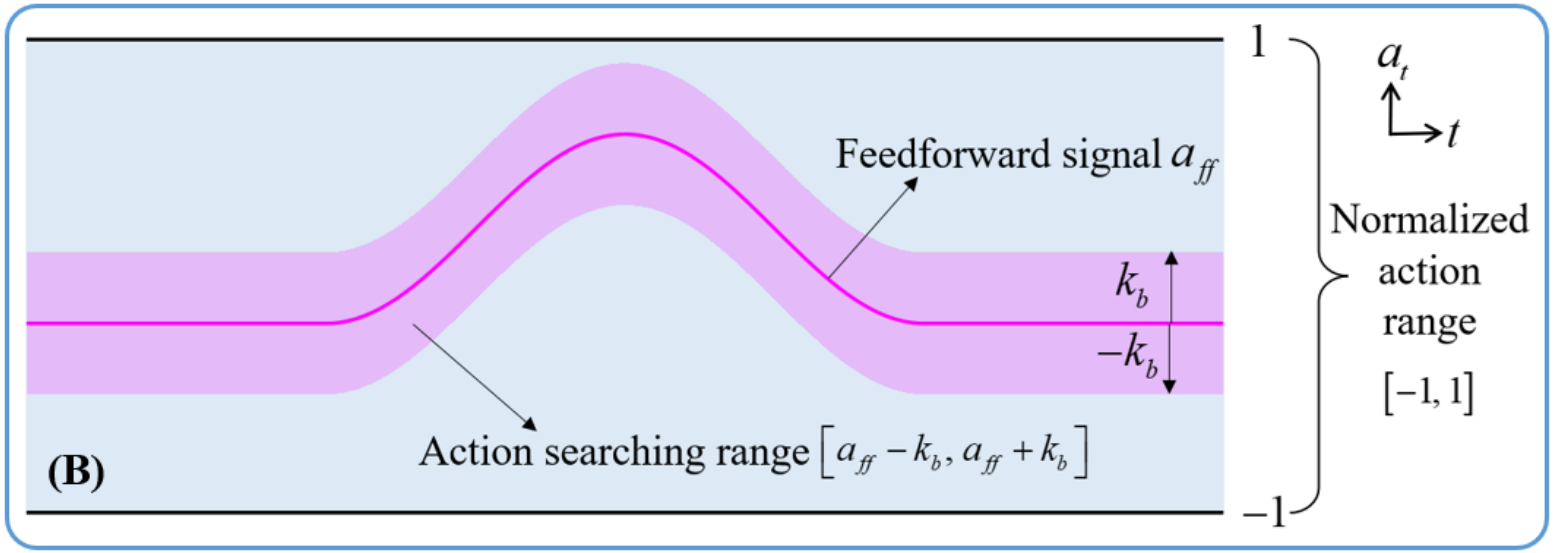
Mimic reward for trajectory tracking

Xue Bin Peng 2020, Learning Agile Robotic Locomotion Skills by Imitating Animals

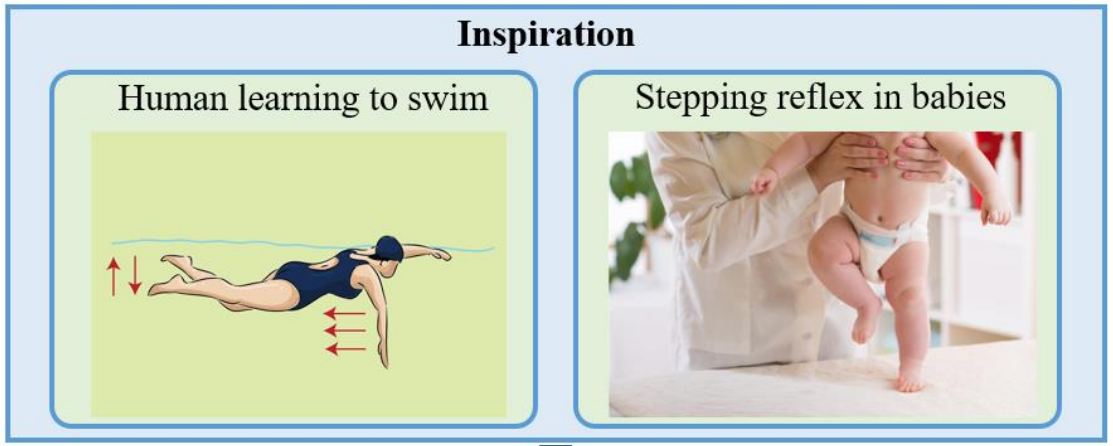
Instruction Learning Framework

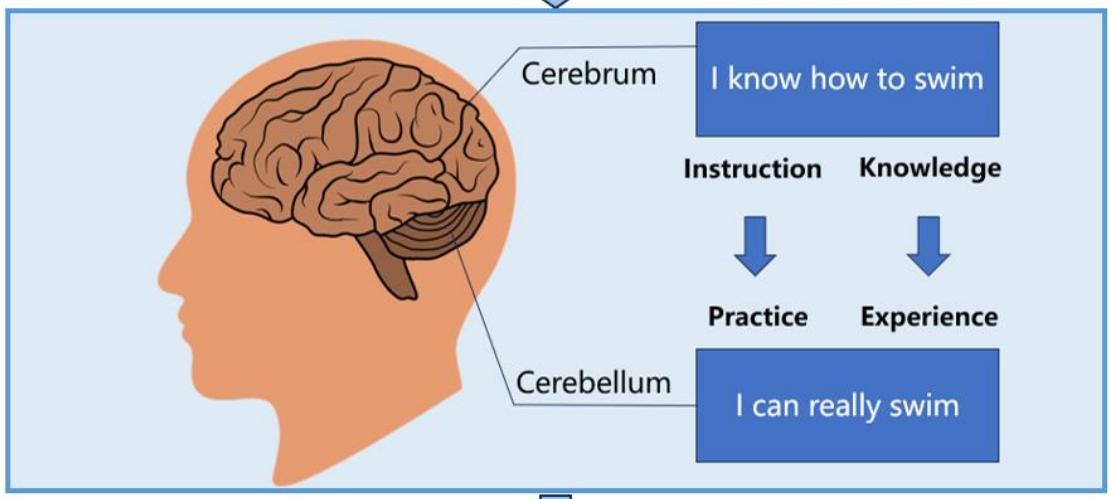


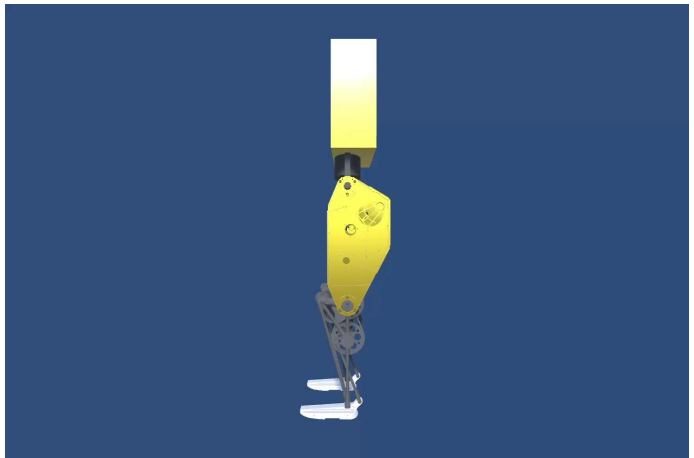
Action Bounding Technique

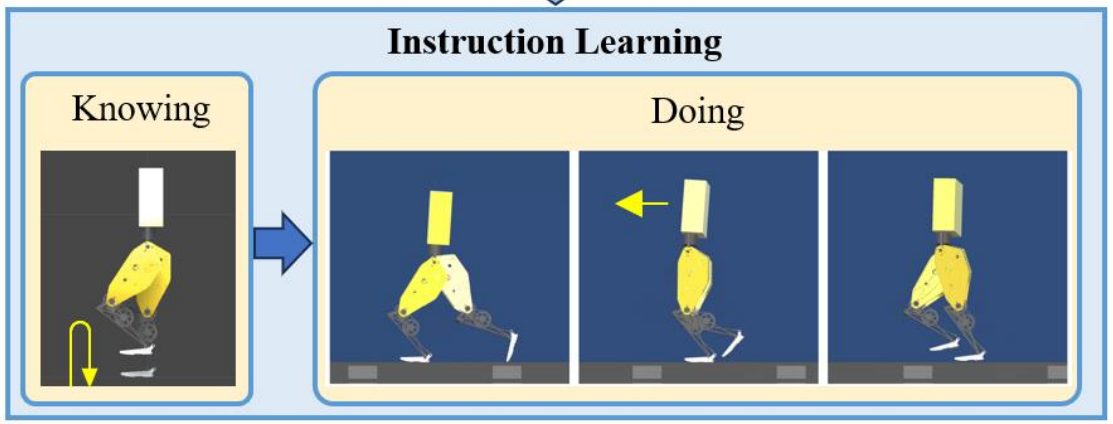


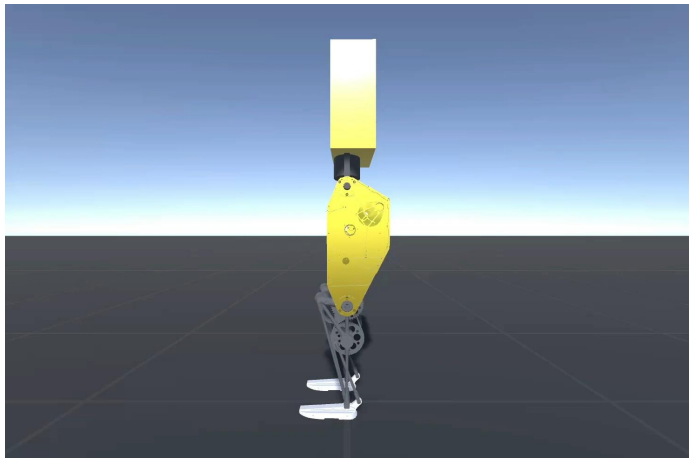
Inspiration from human learning









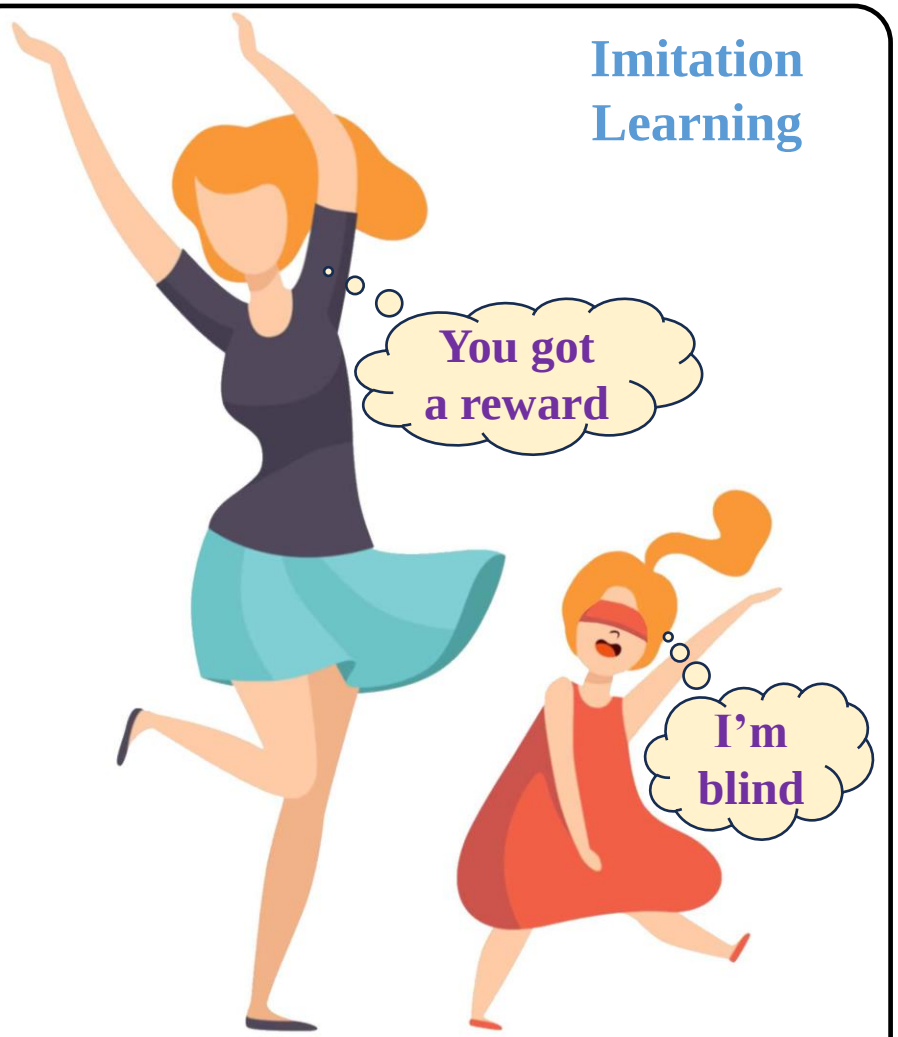


Imitation Learning

vs.

Instruction Learning

Imitation Learning



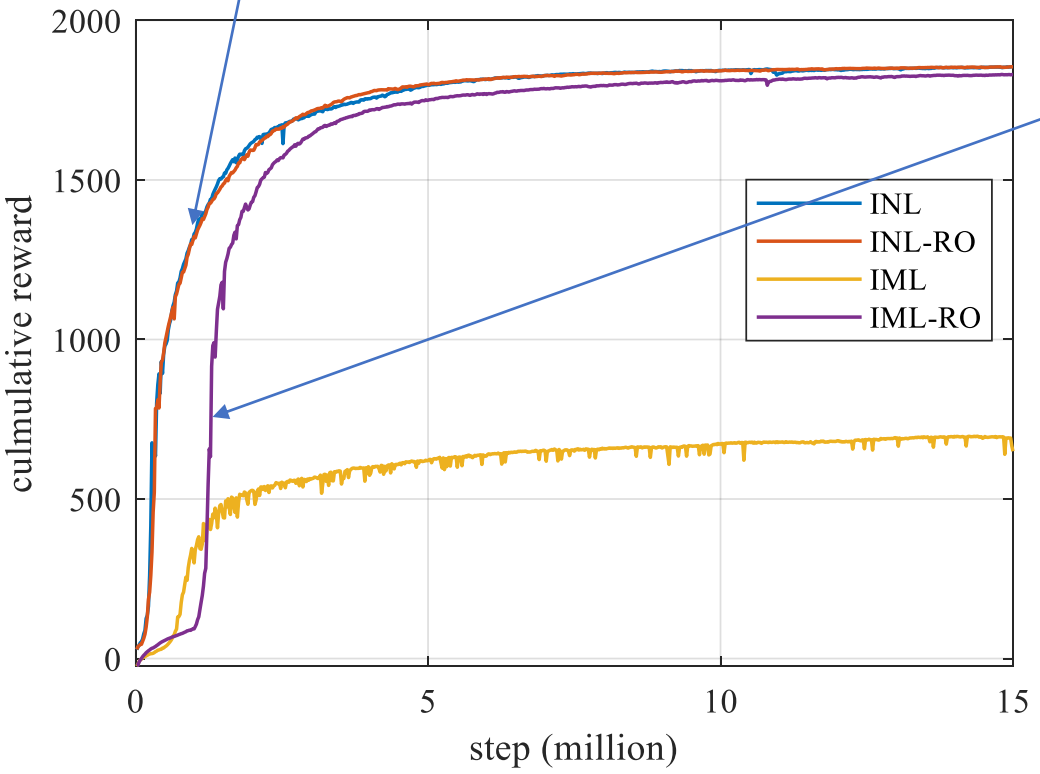
- Learn from scratch
- Full action space
- Purely reward-driven
- Require mimic reward

Instruction Learning

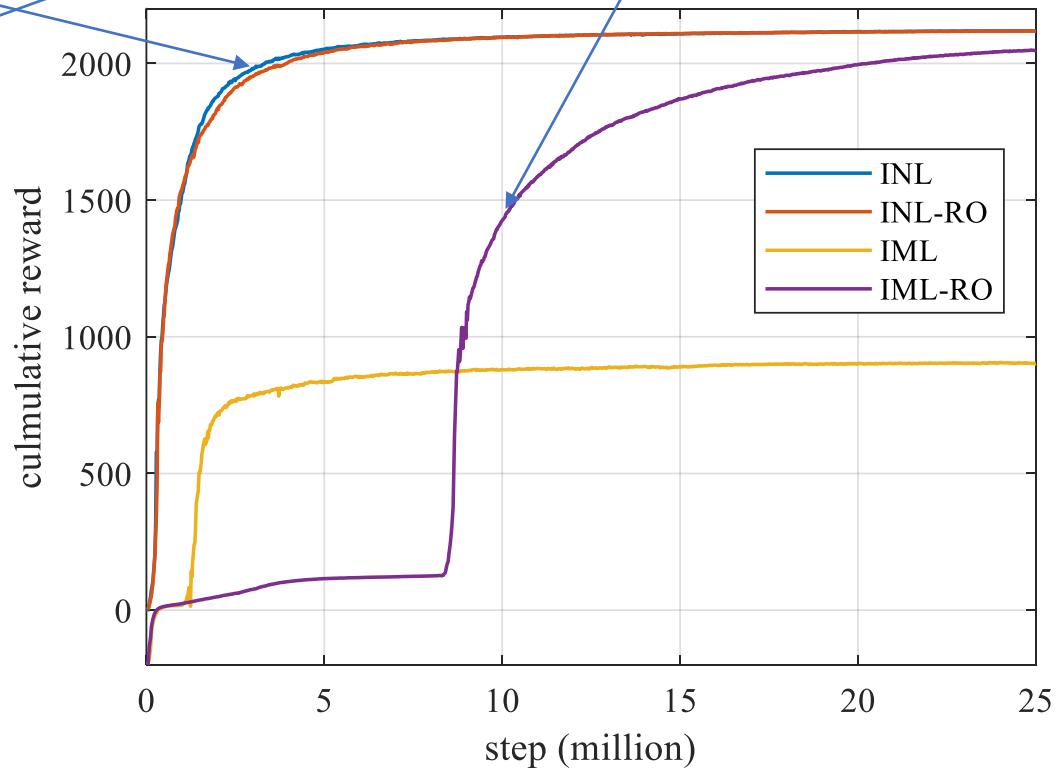
- Learn by following instruction
- Reduced action space
- Reward and feedforward driven
- No need for mimic reward

Instruction learning (INL) learns much faster than imitation learning (IML)

Biped

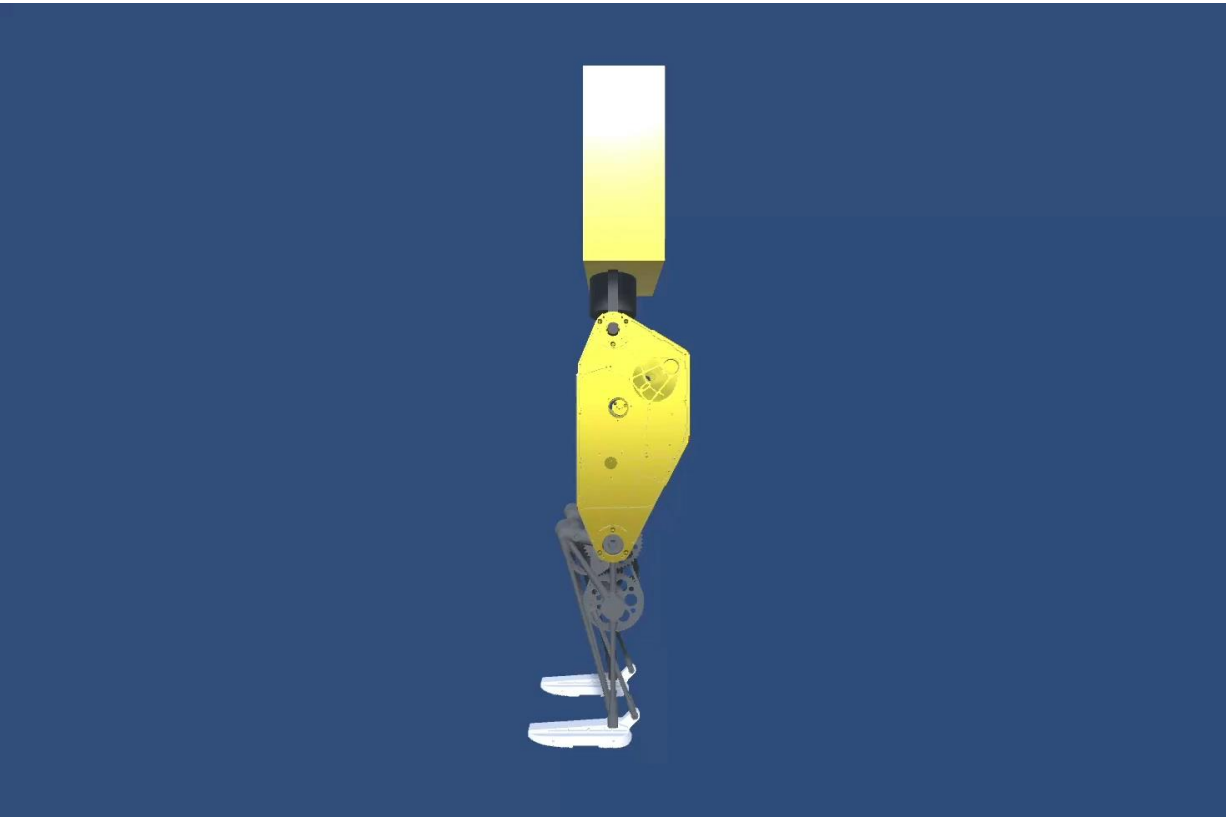


Quadruped

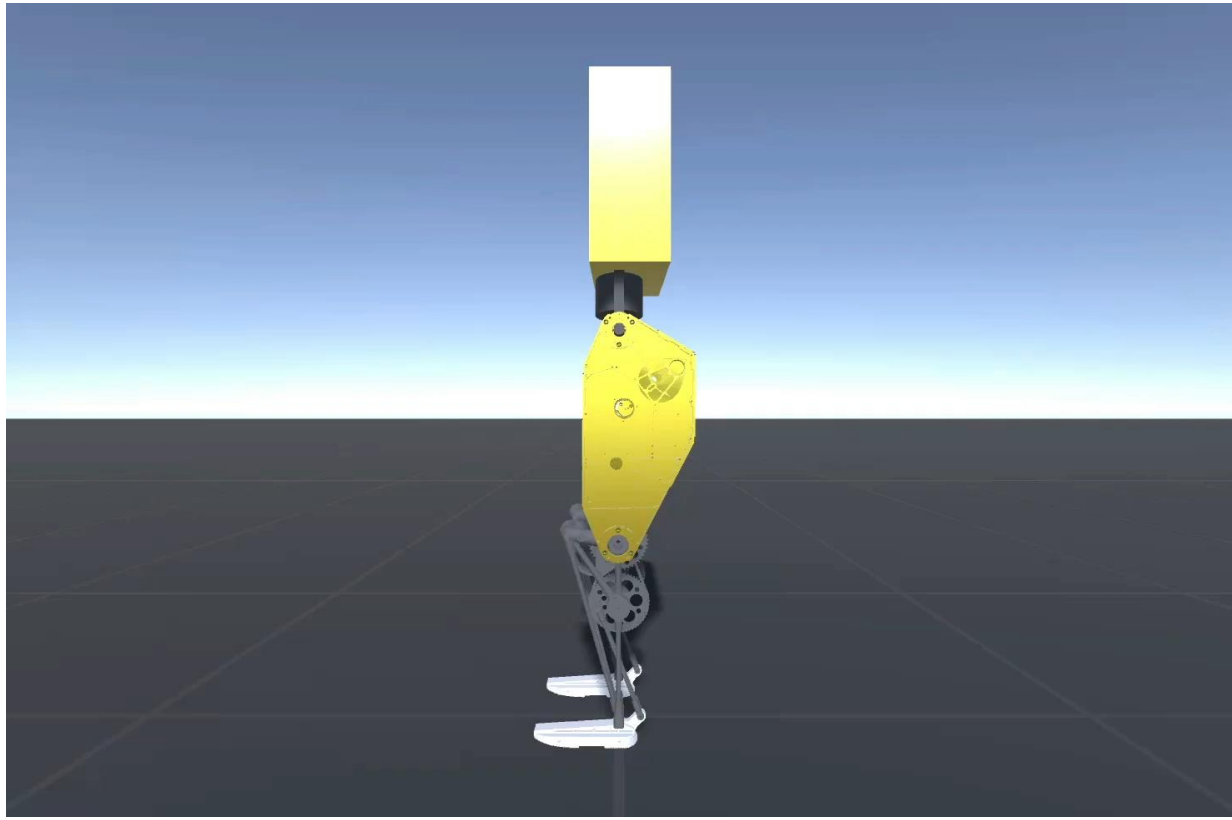


Biped Stepping

Feedforward action

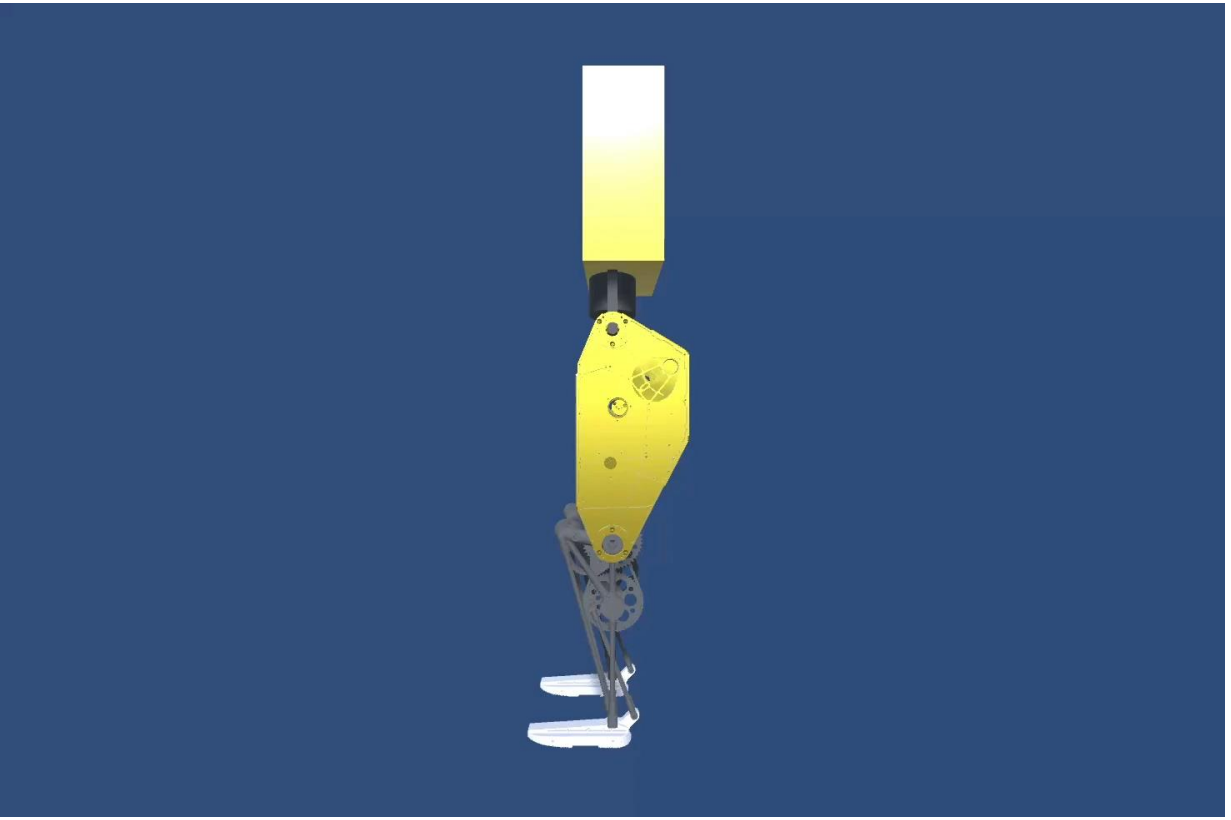


Learned motion

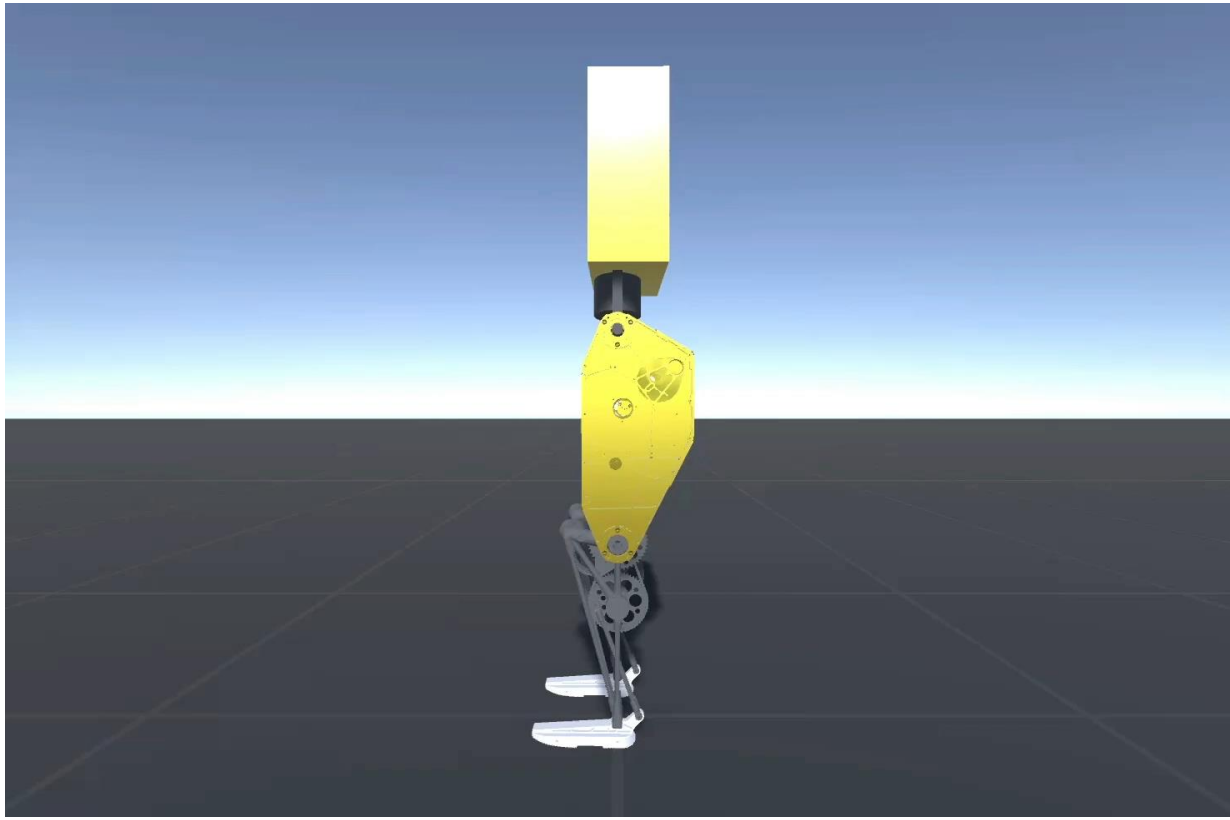


Biped Walking

Feedforward action

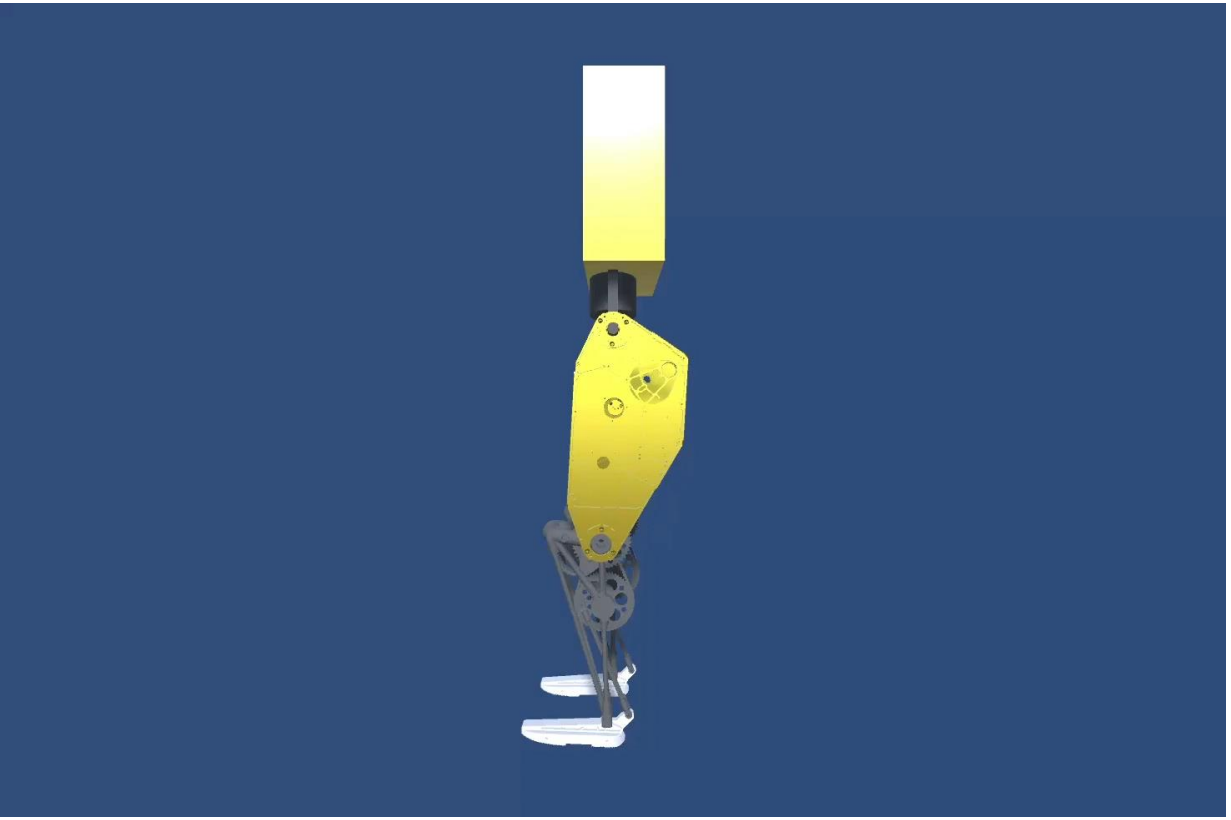


Learned motion

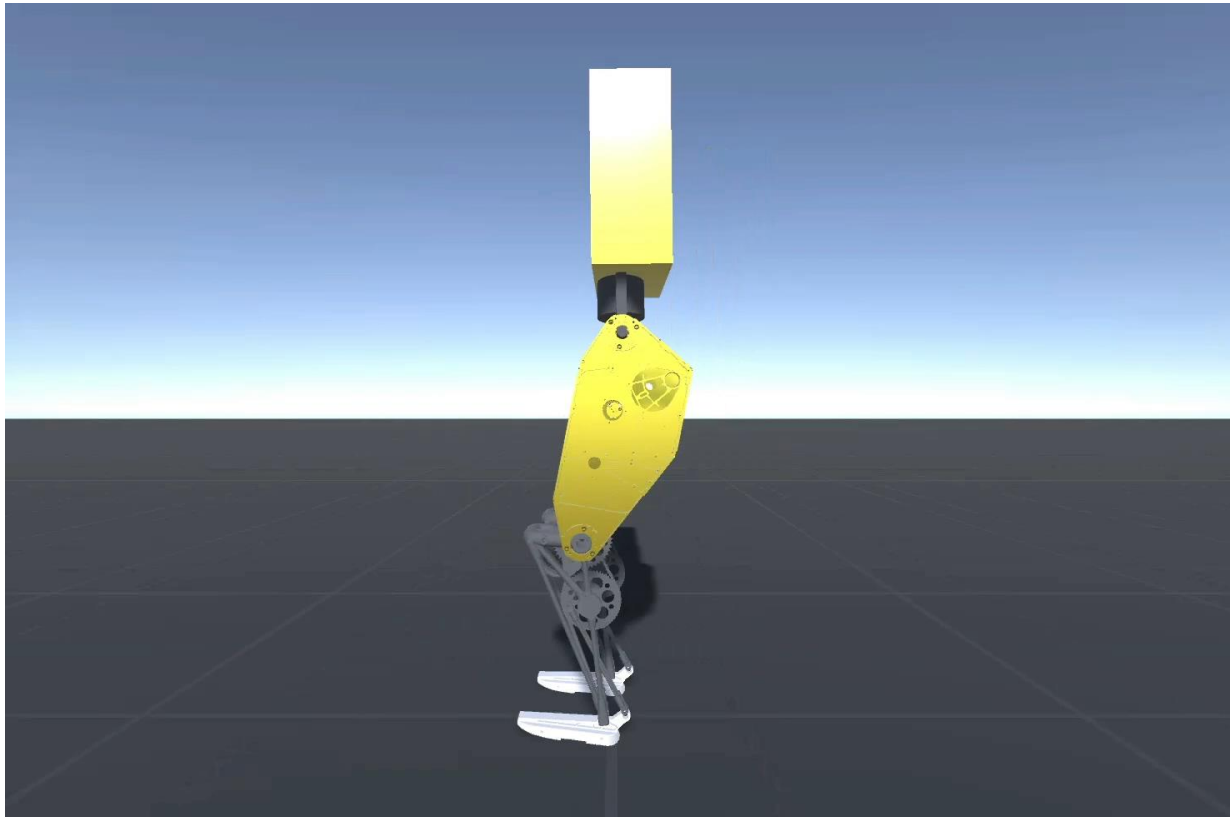


Biped Level Walking

Feedforward action

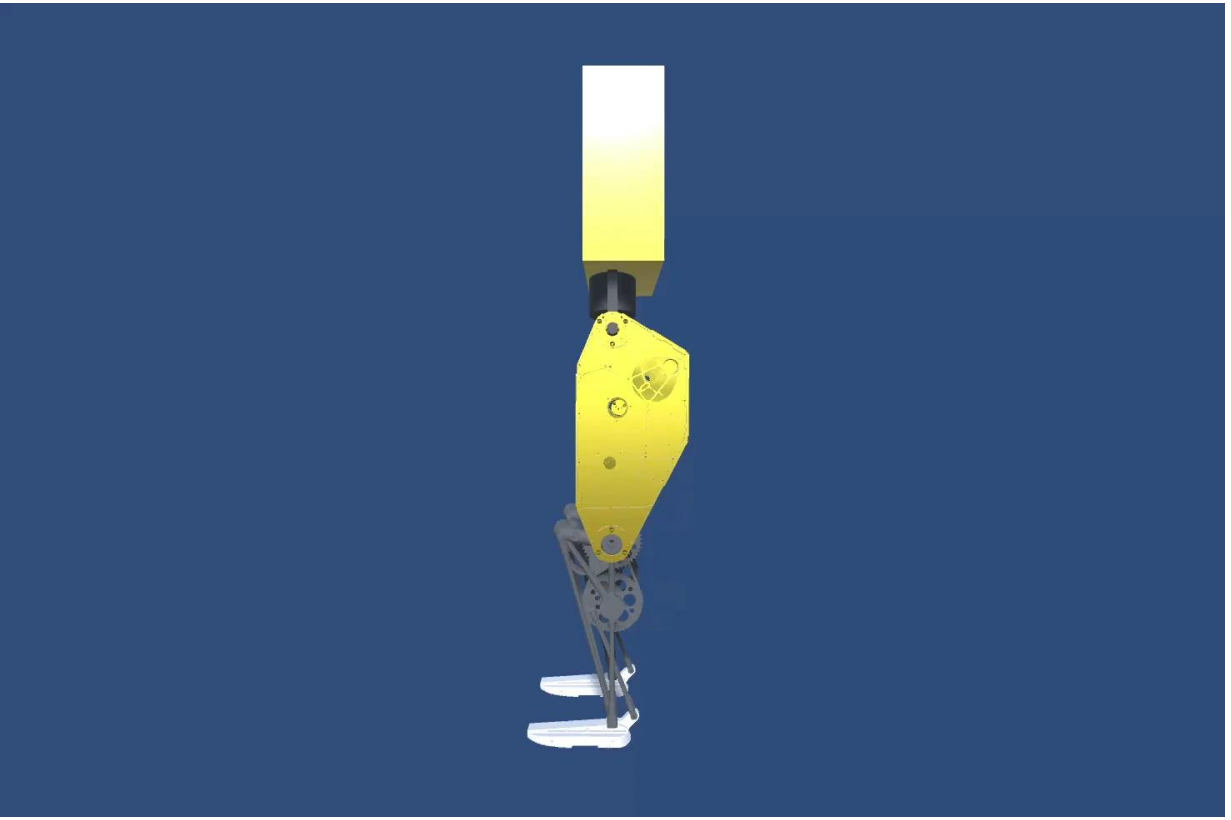


Learned motion

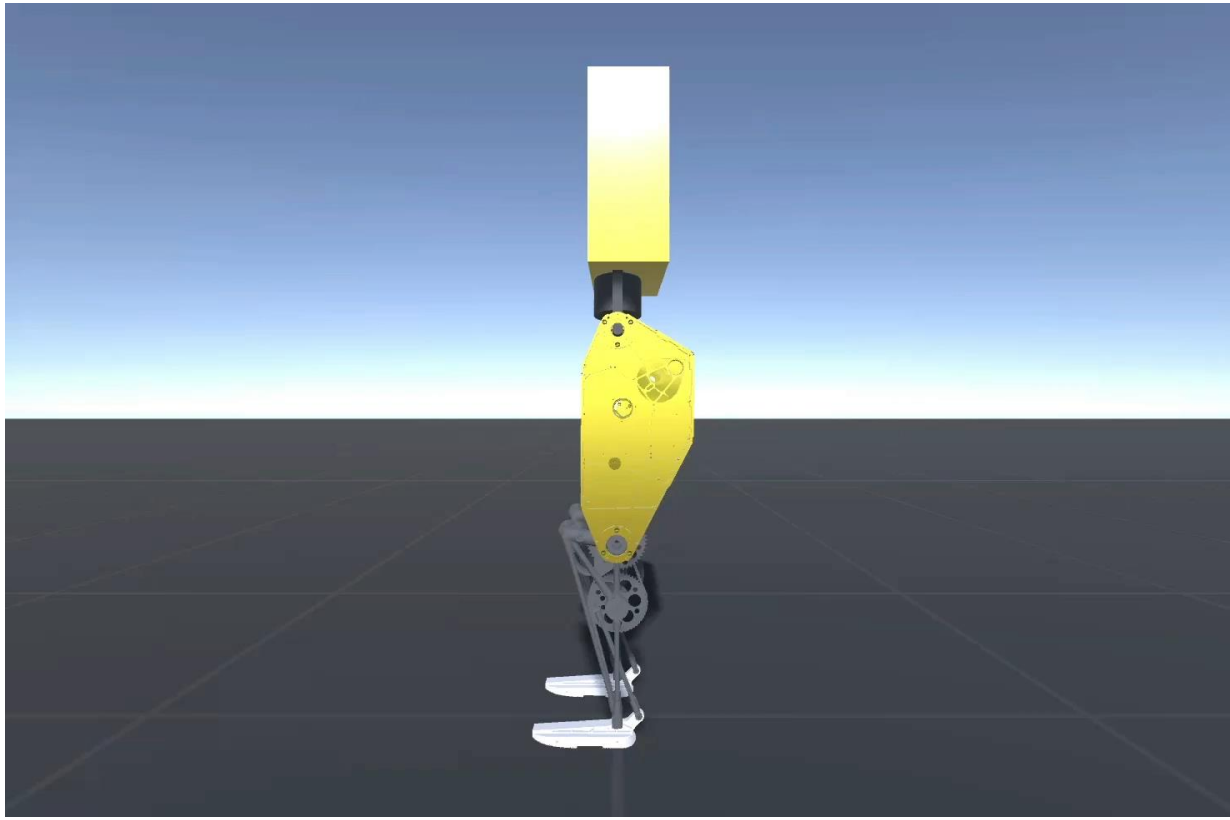


Biped March Walking

Feedforward action

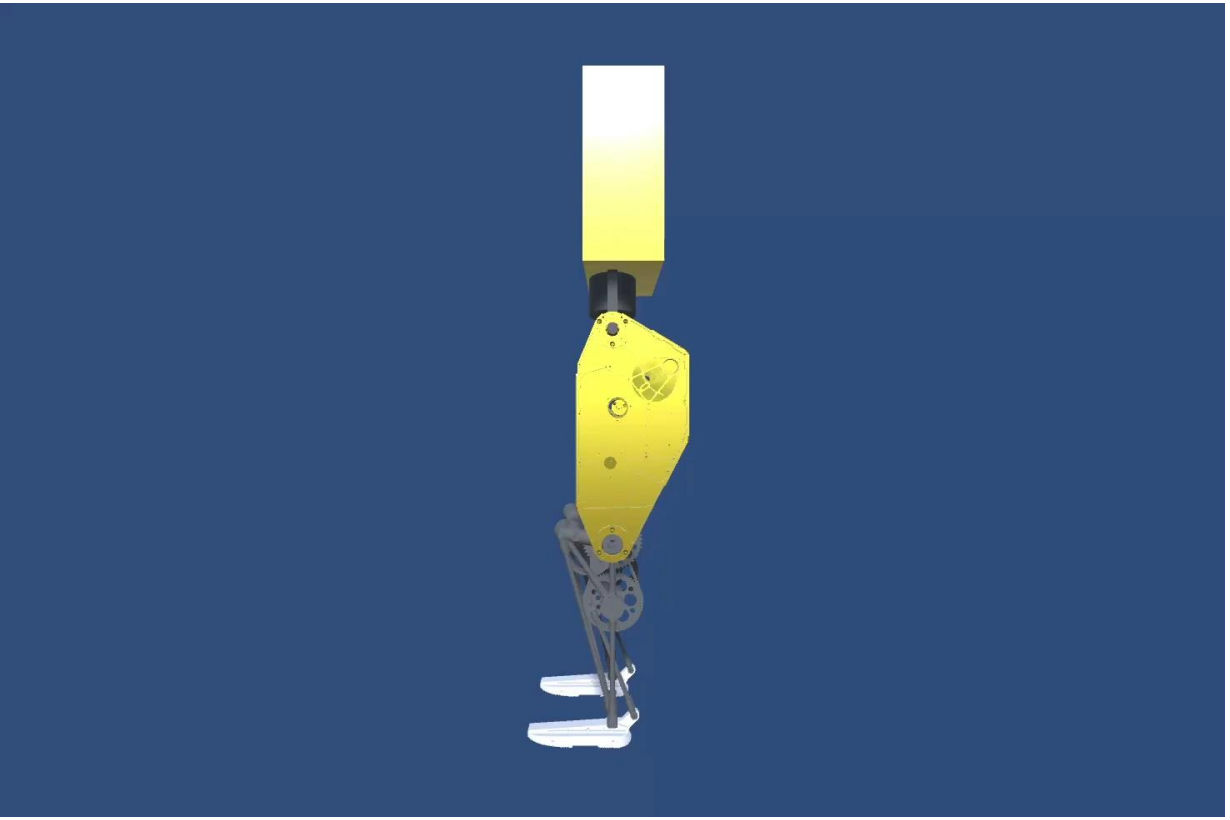


Learned motion

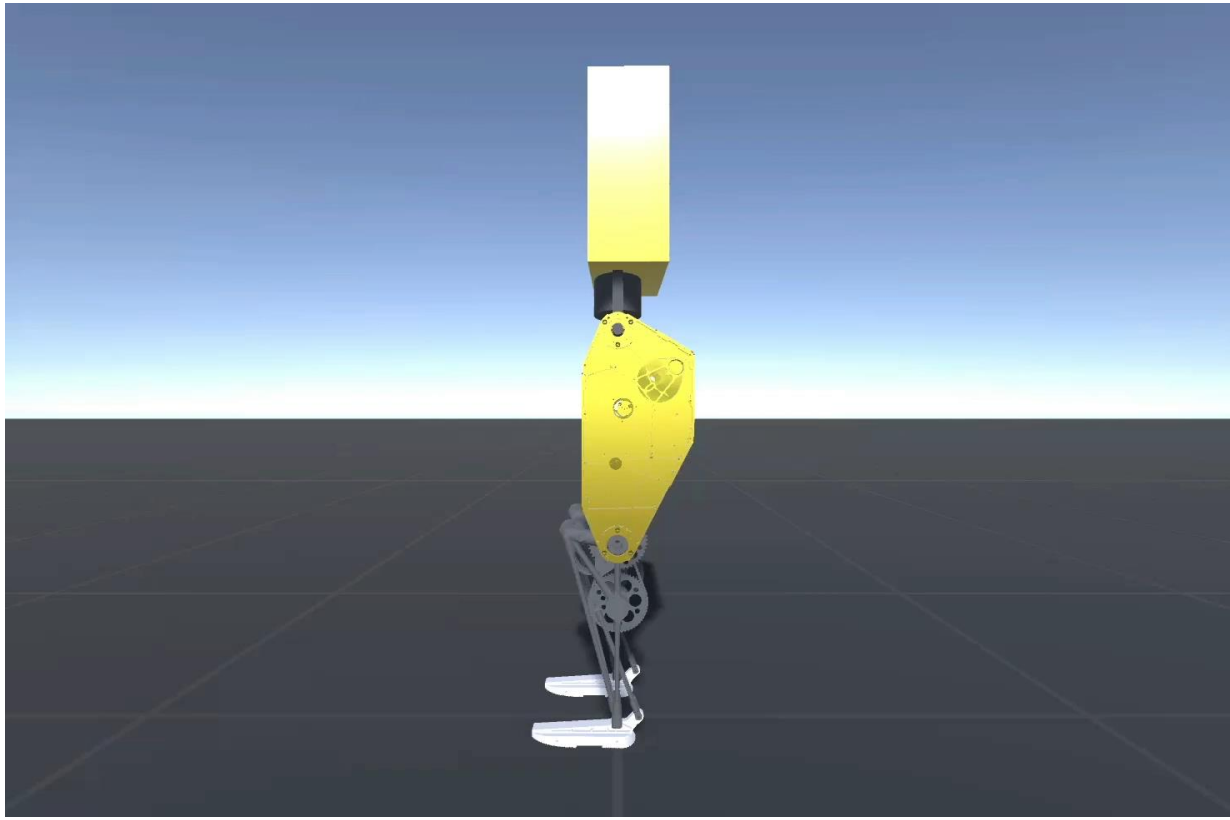


Biped Jumping

Feedforward action

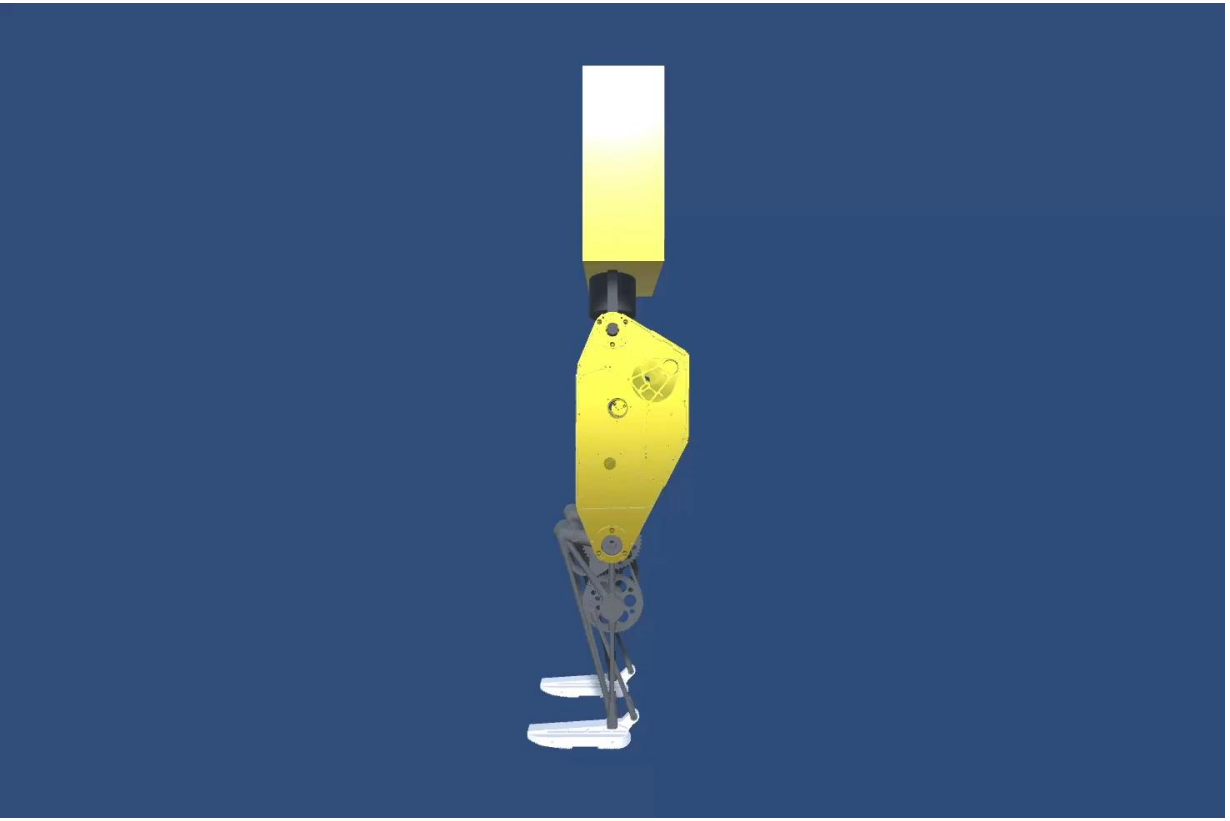


Learned motion

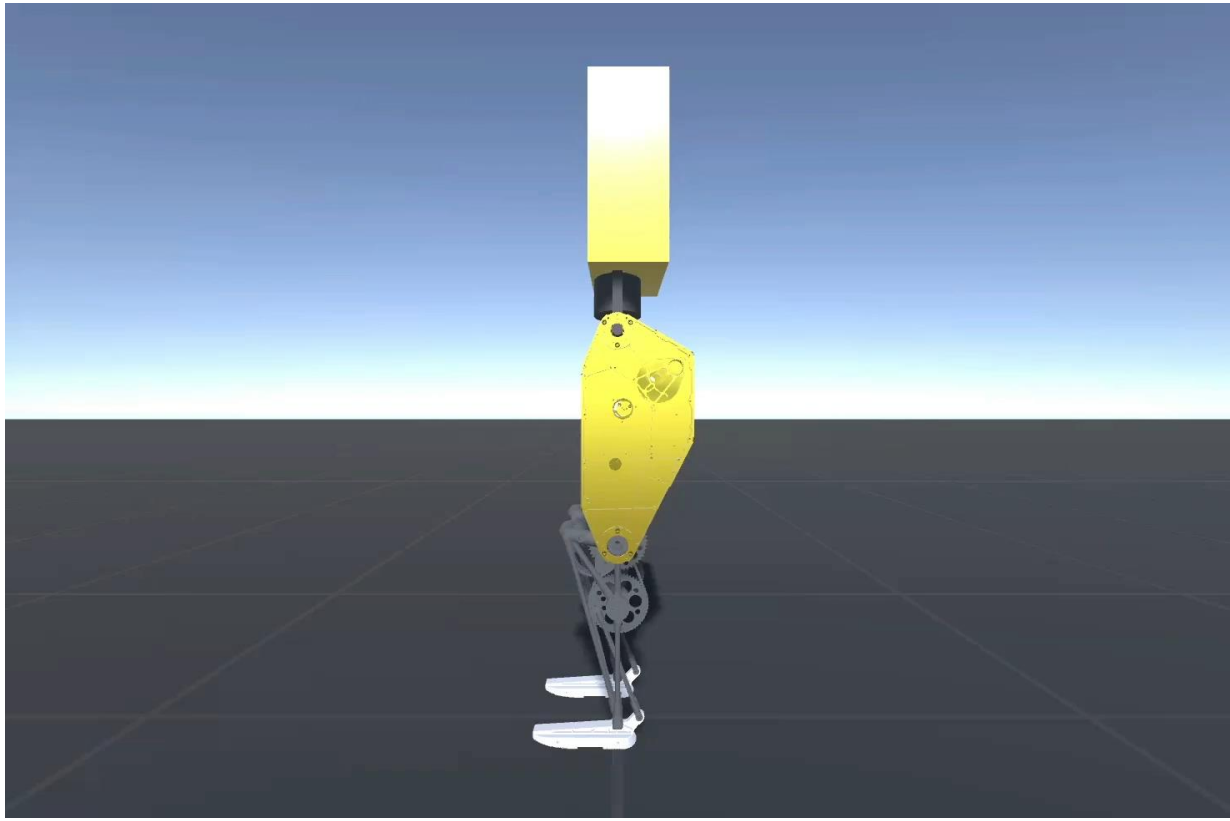


Biped Hopping

Feedforward action



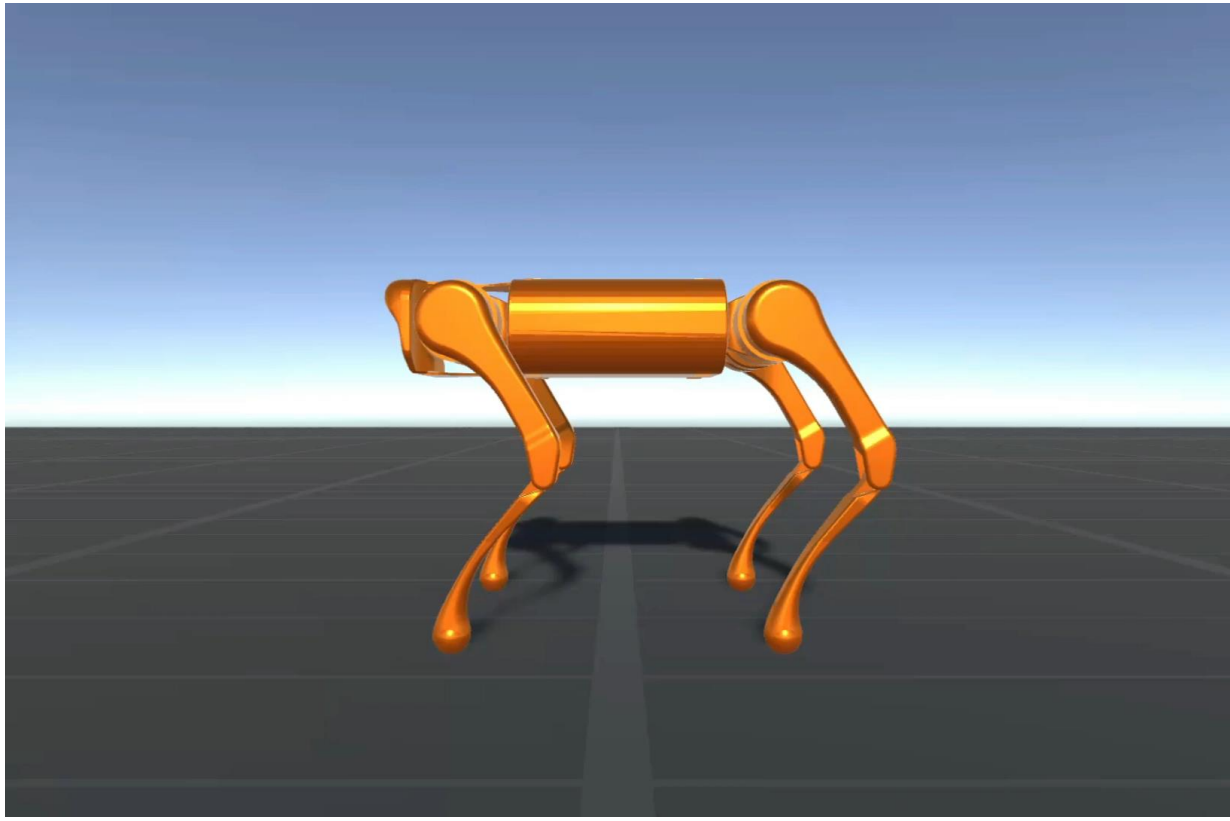
Learned motion



Quadruped Stepping

Feedforward action

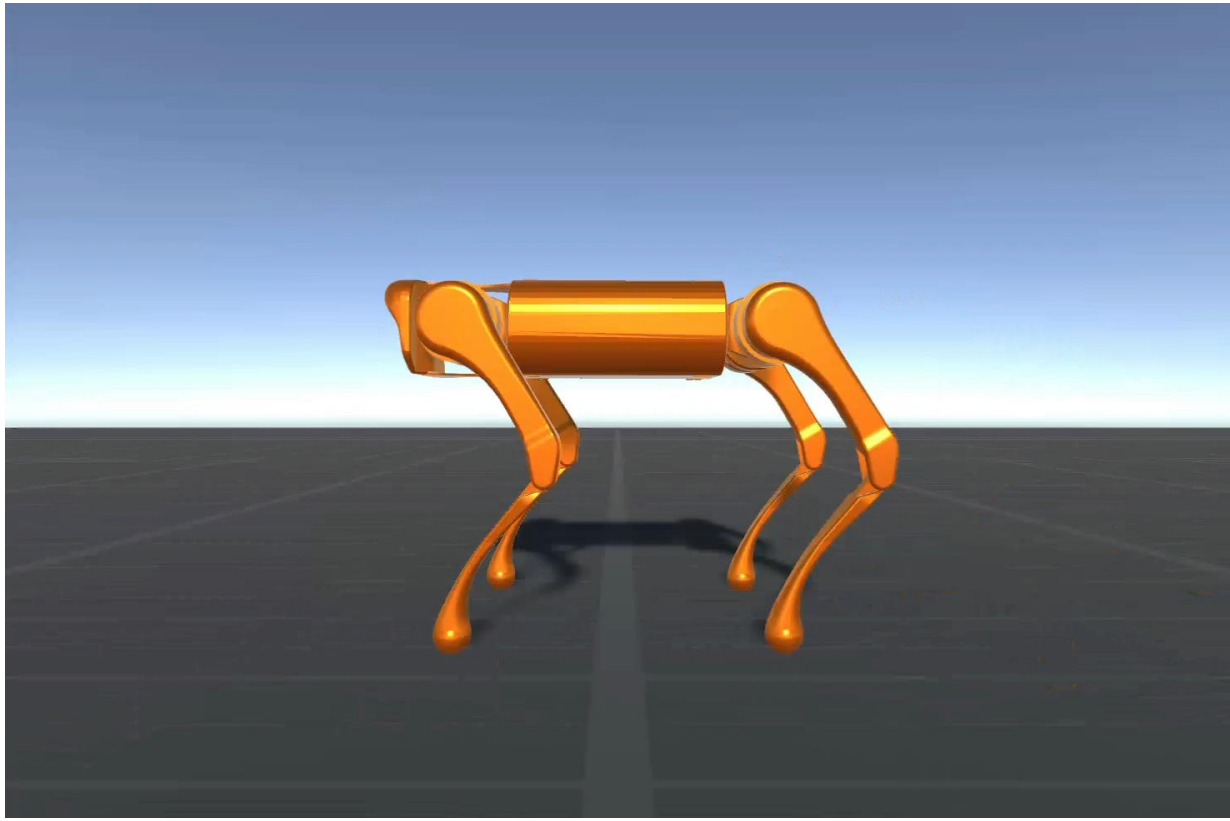
Learned motion



Quadruped Trotting

Feedforward action

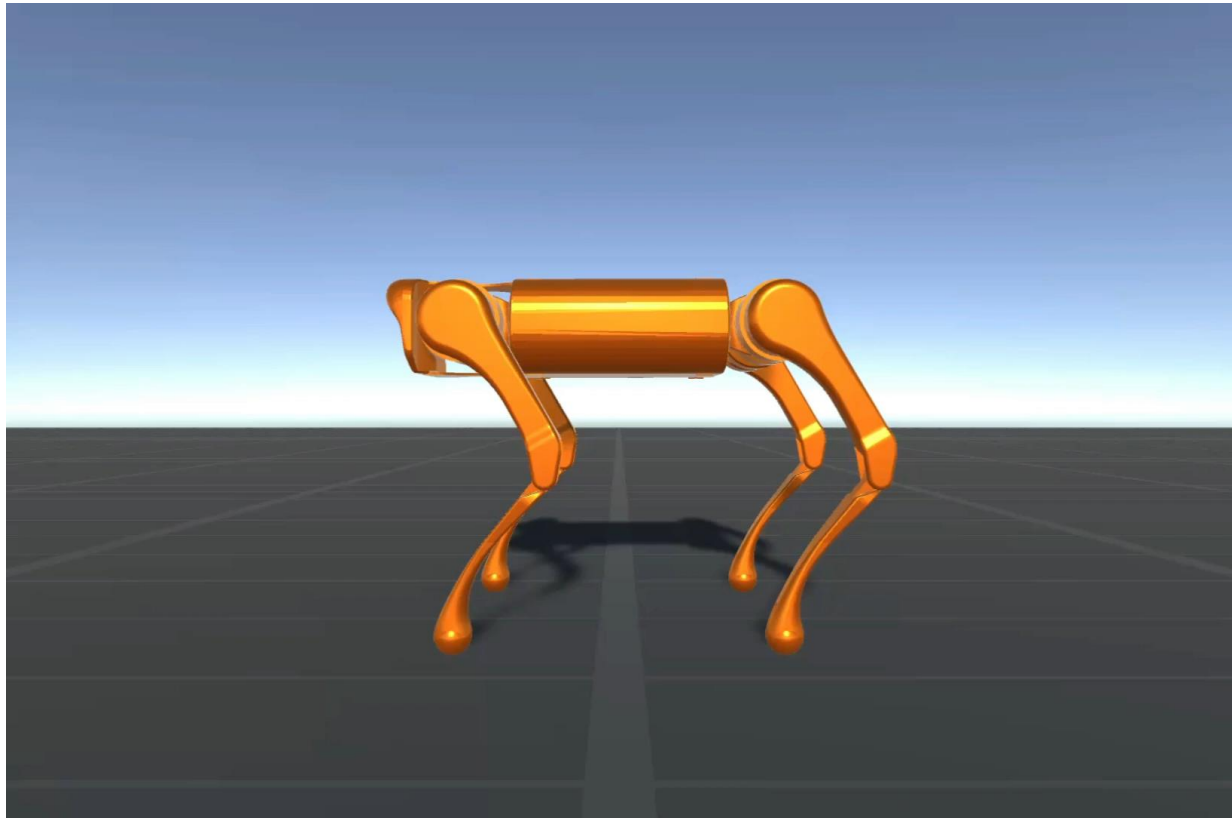
Learned motion



Quadruped Pacing

Feedforward action

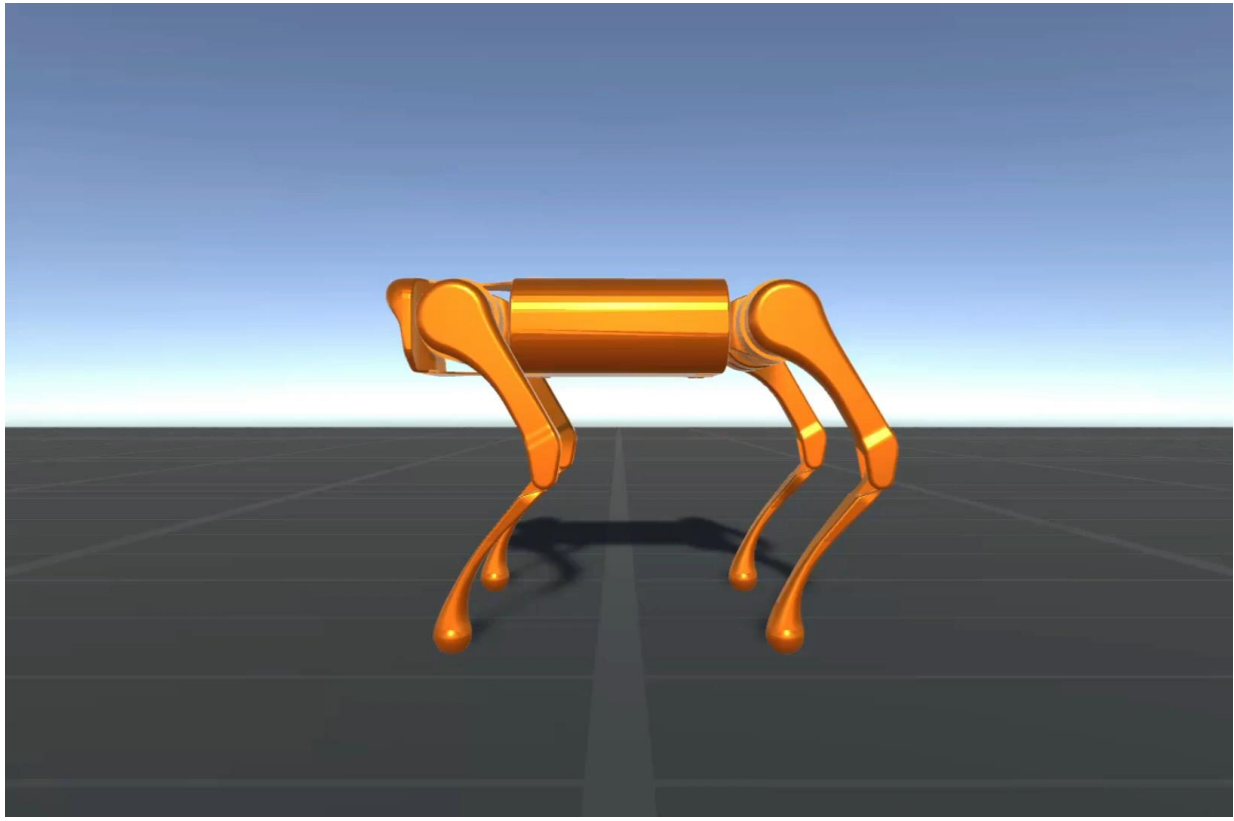
Learned motion



Quadruped Bounding

Feedforward action

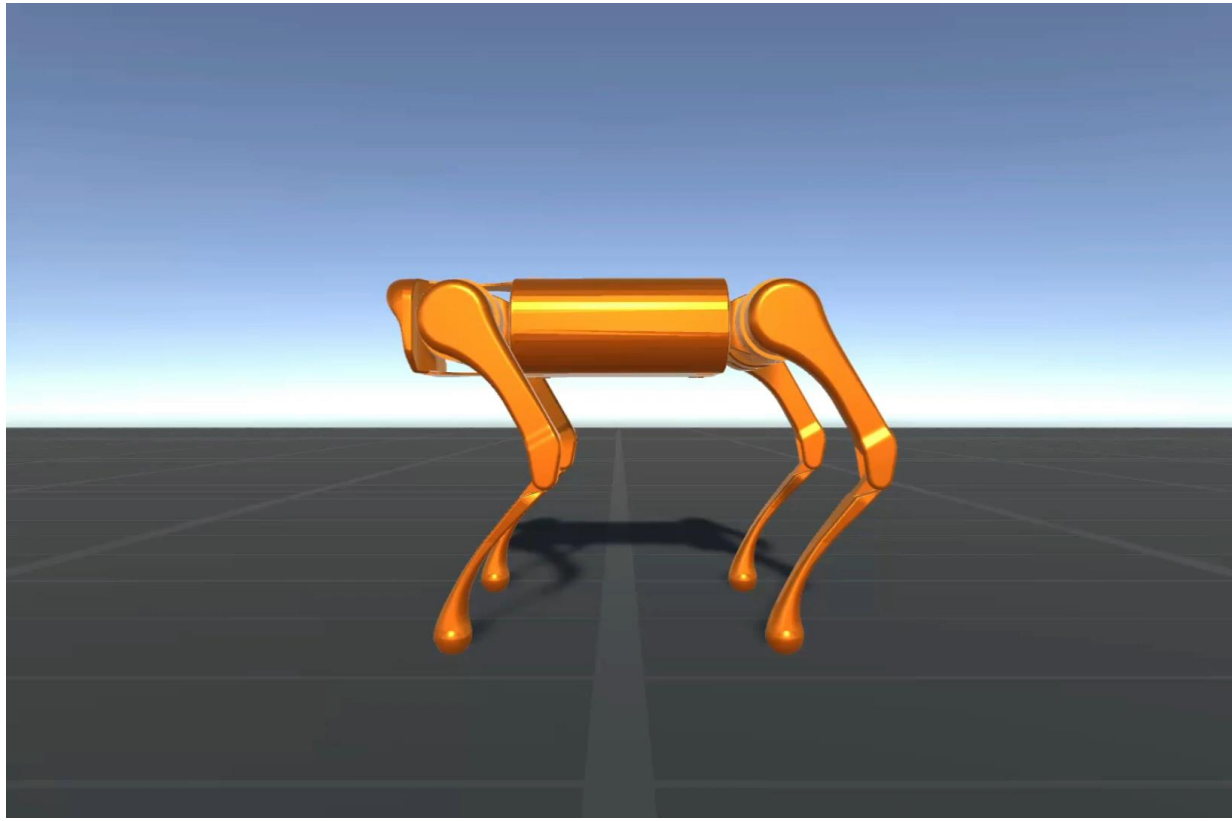
Learned motion



Quadruped Pronking

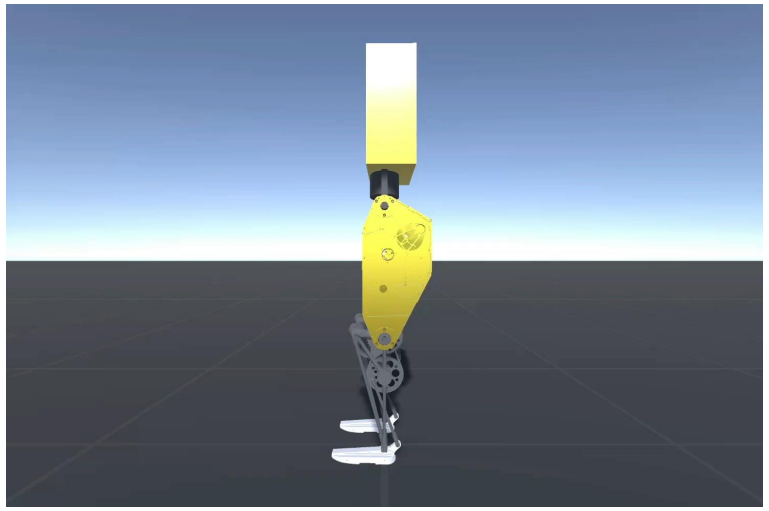
Feedforward action

Learned motion

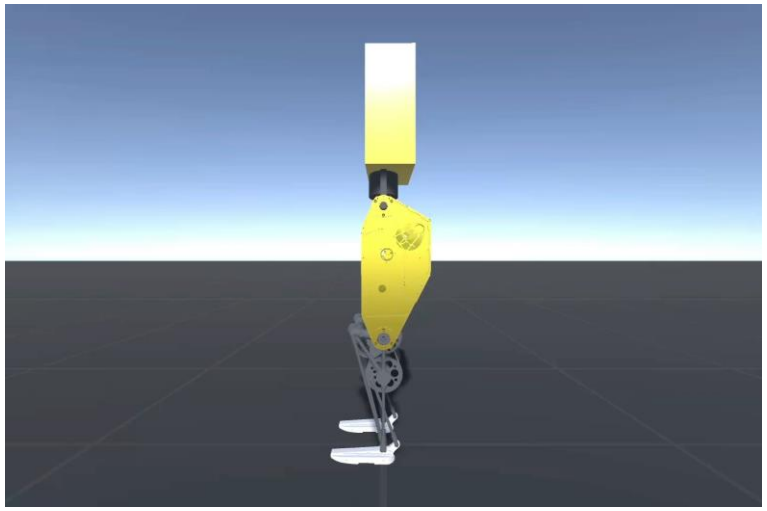


Motion Adaption

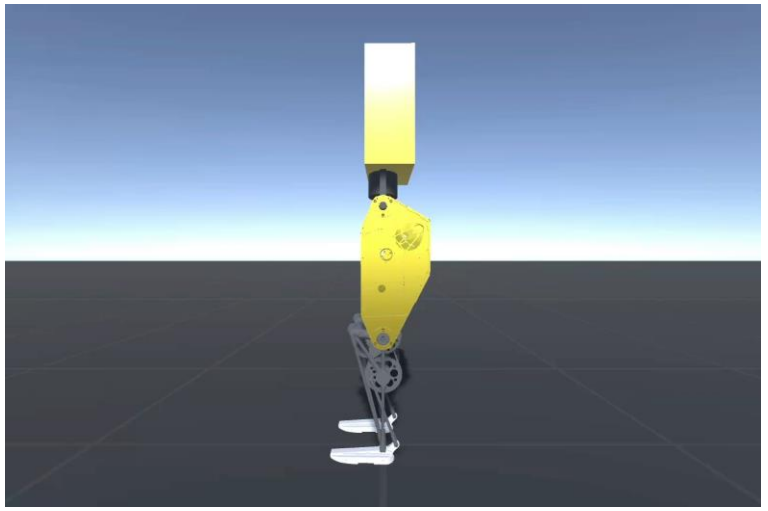
Learned motion



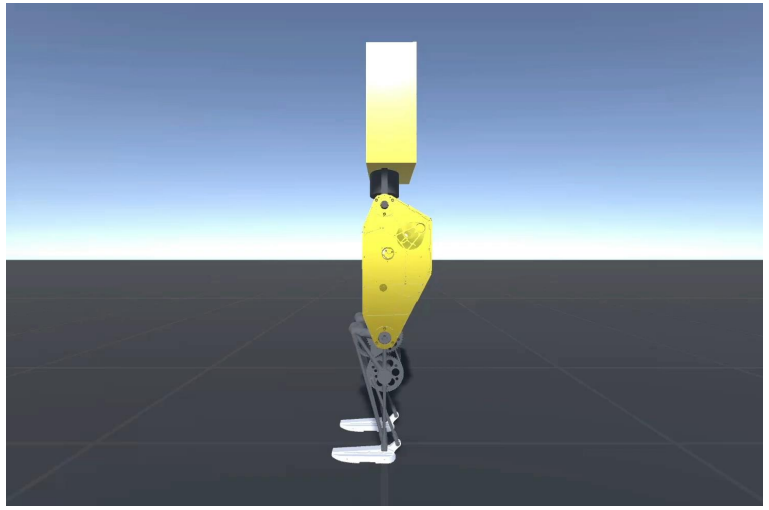
Period * 0.88



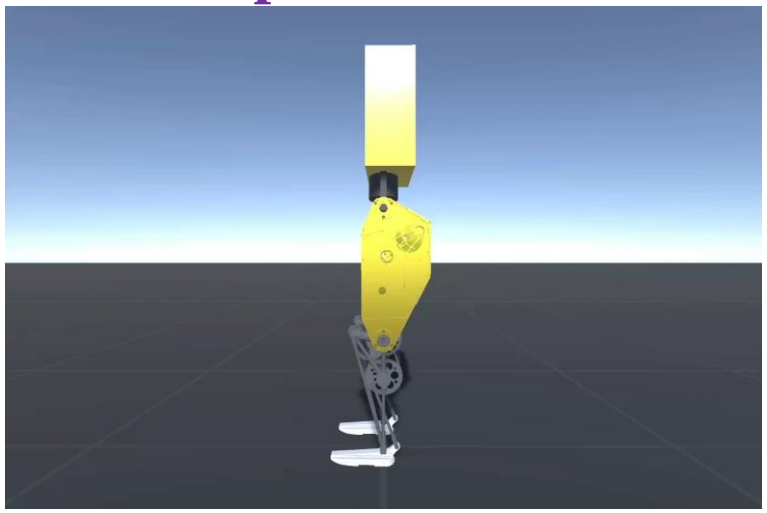
Period * 1.26



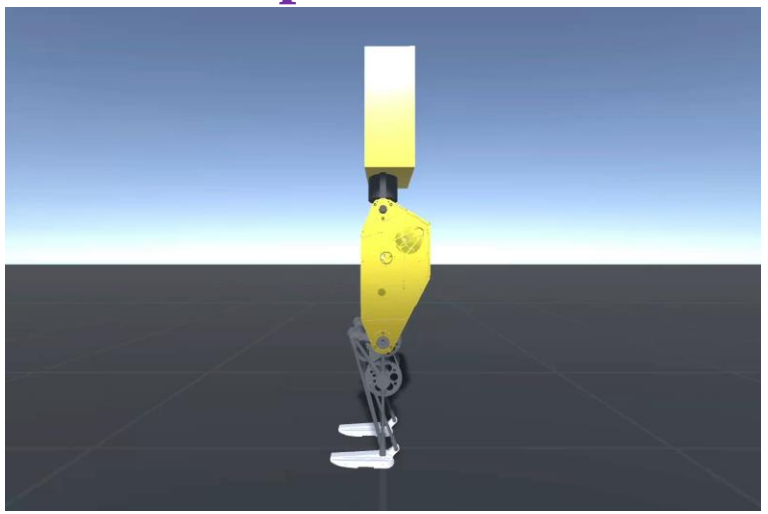
Learned motion



Amplitude * 0.88

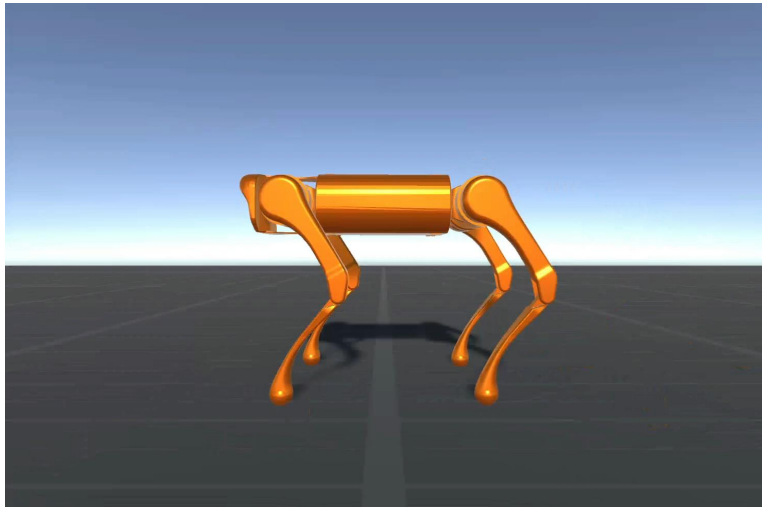


Amplitude * 2.08

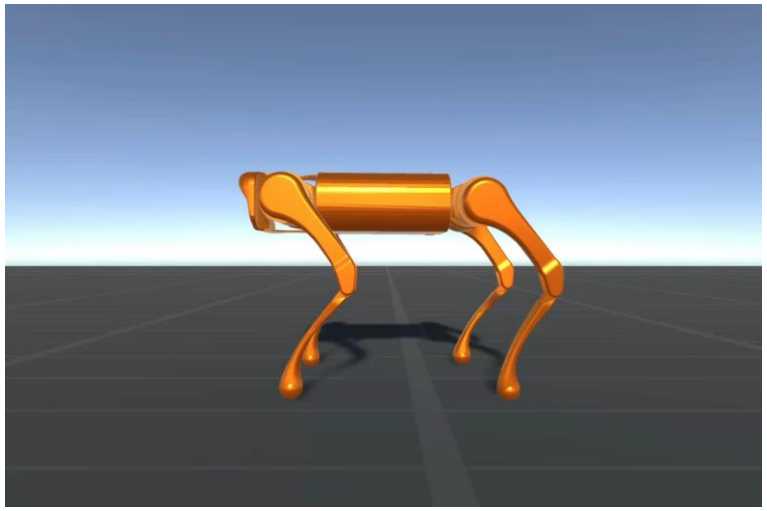


Motion Adaption

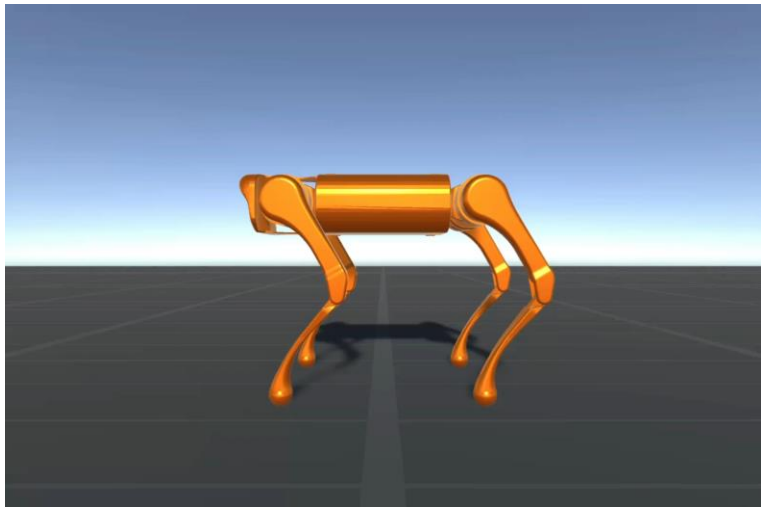
Learned motion



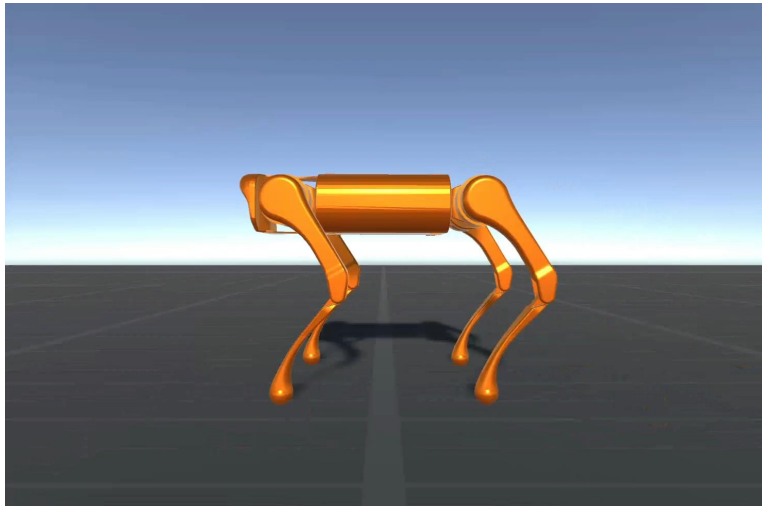
Period * 0.5



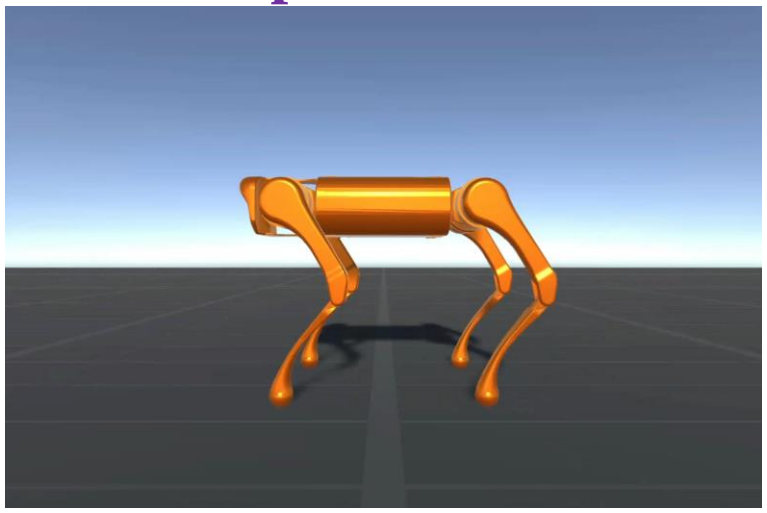
Period * 1.4



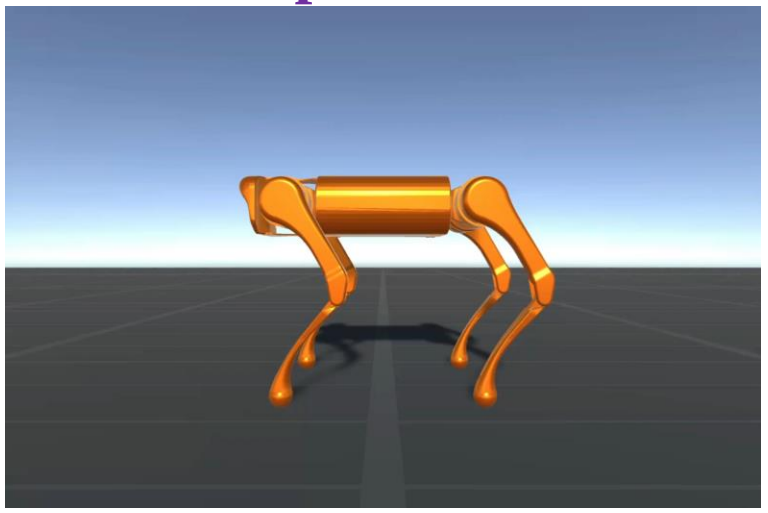
Learned motion



Amplitude * 0.7



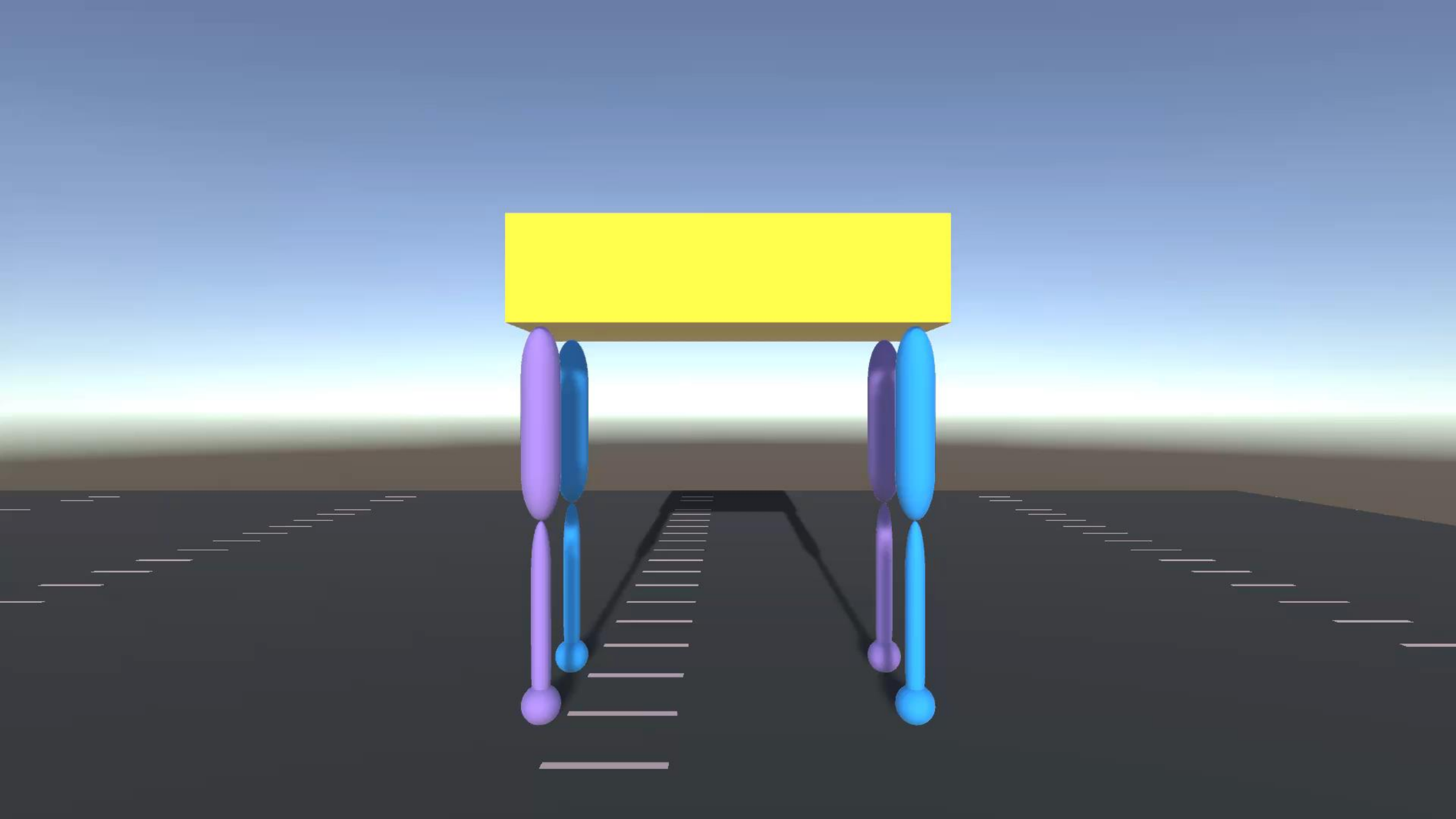
Amplitude * 4.0

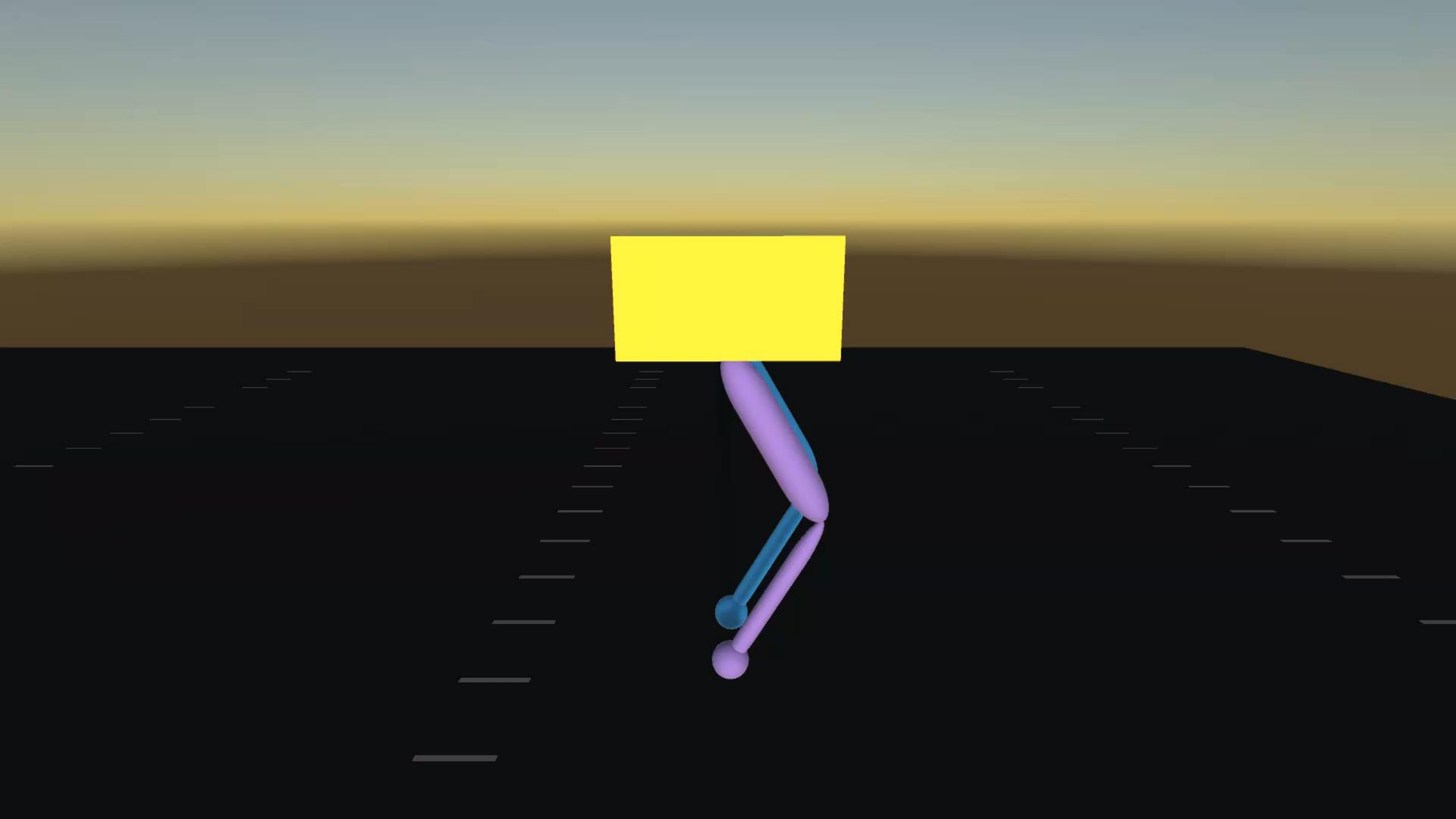


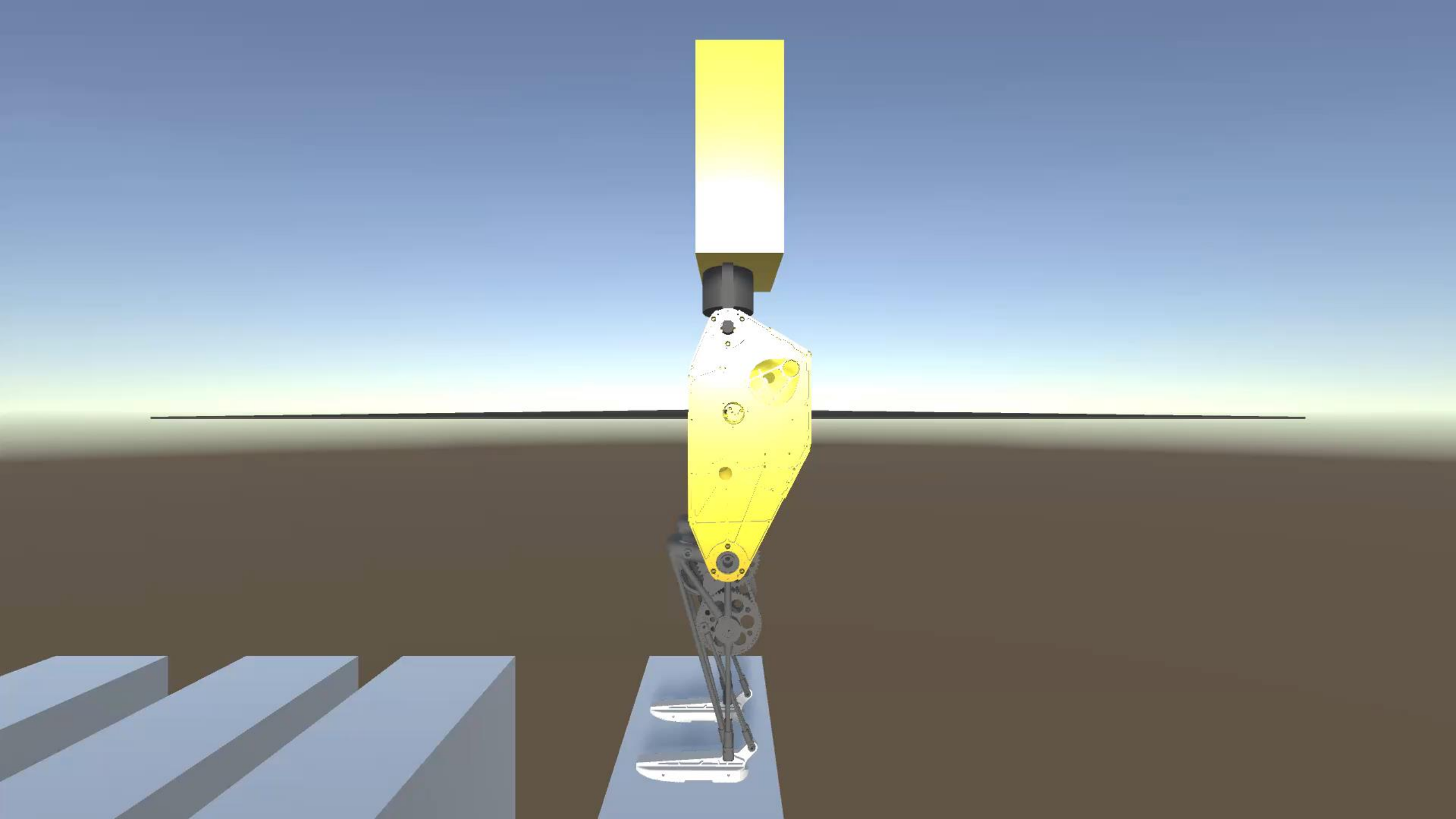
Sim to Real: multiple gaits realization

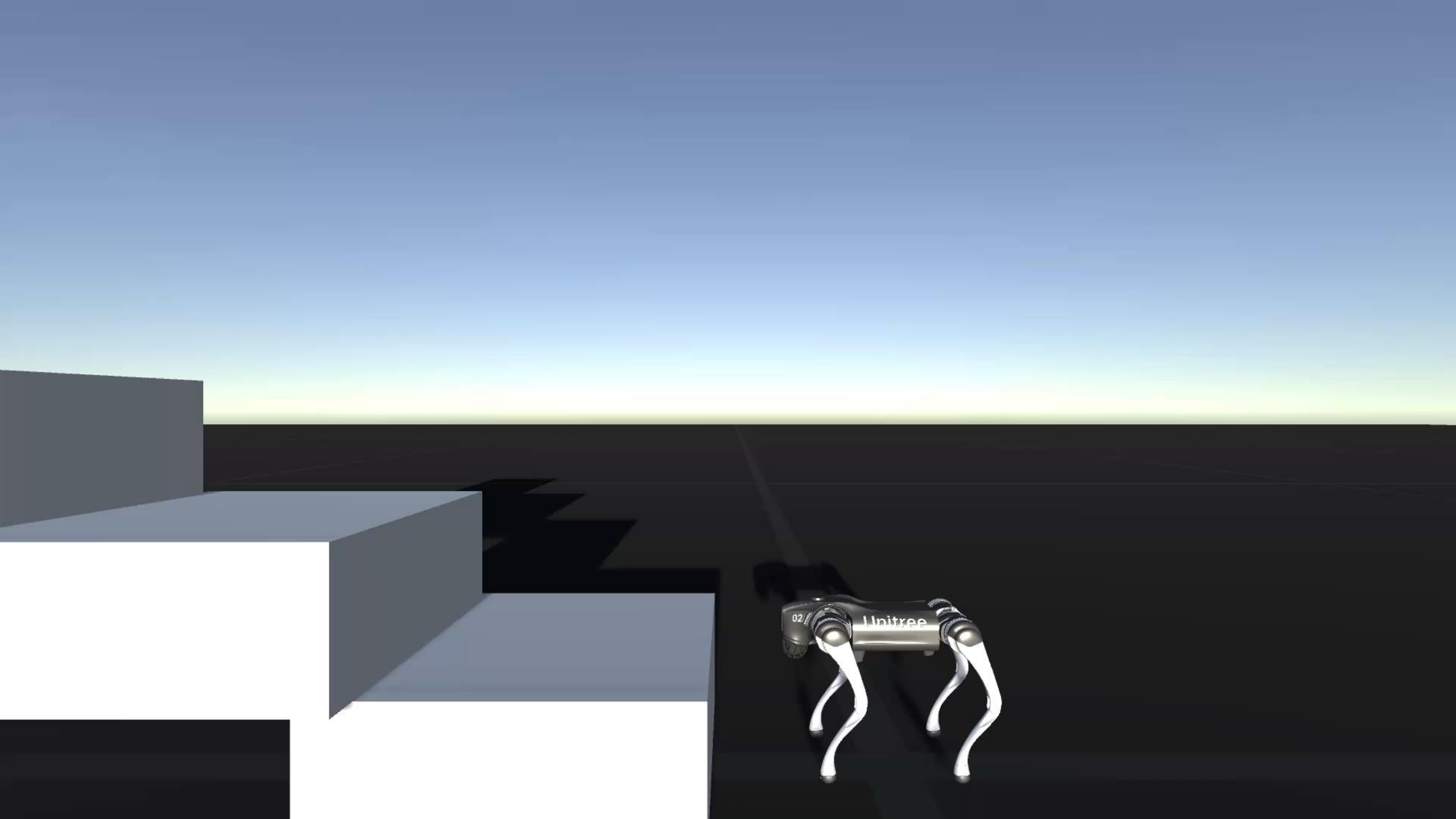
Sim to Real: velocity following and robust walking

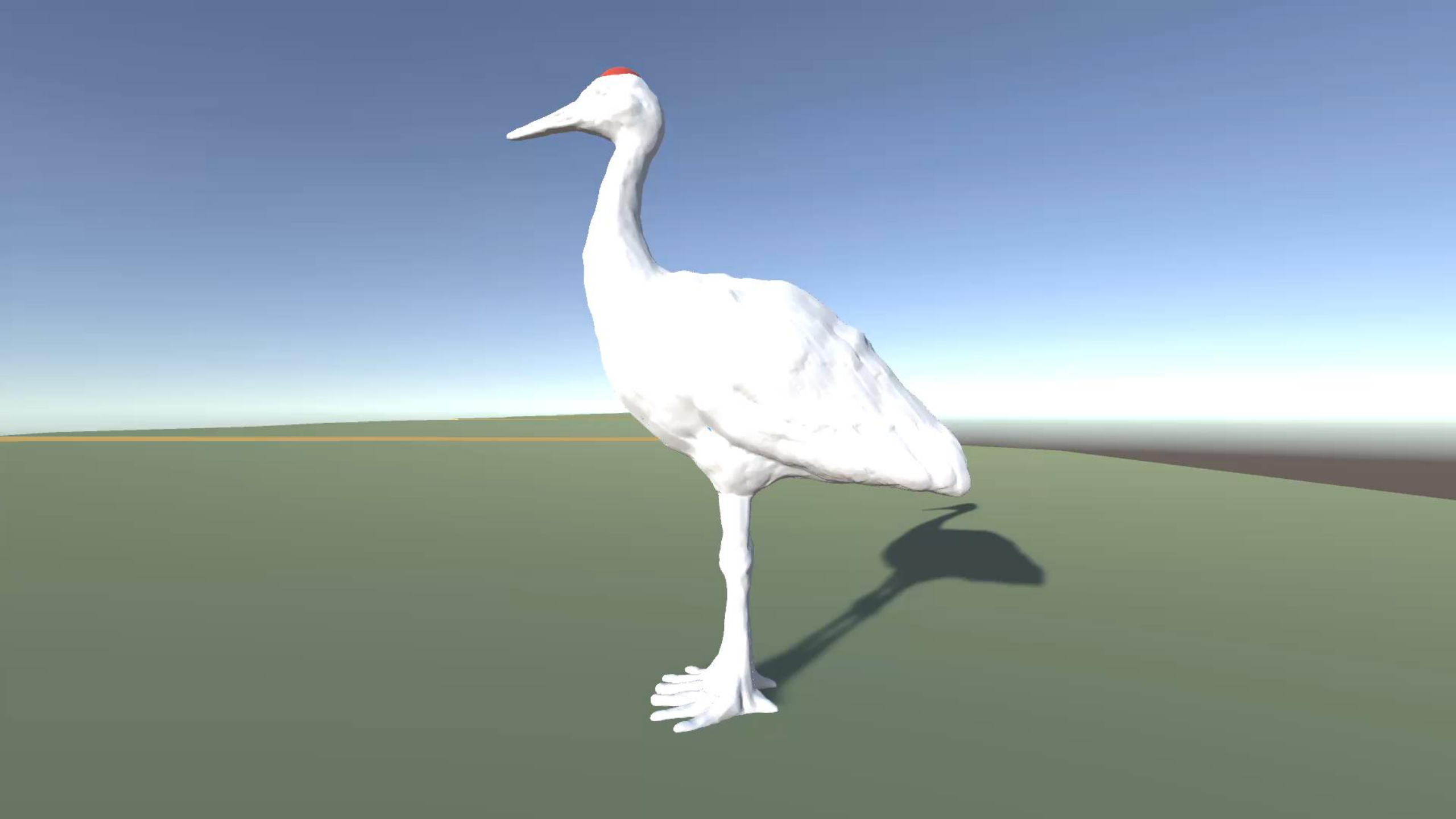
Online Learning

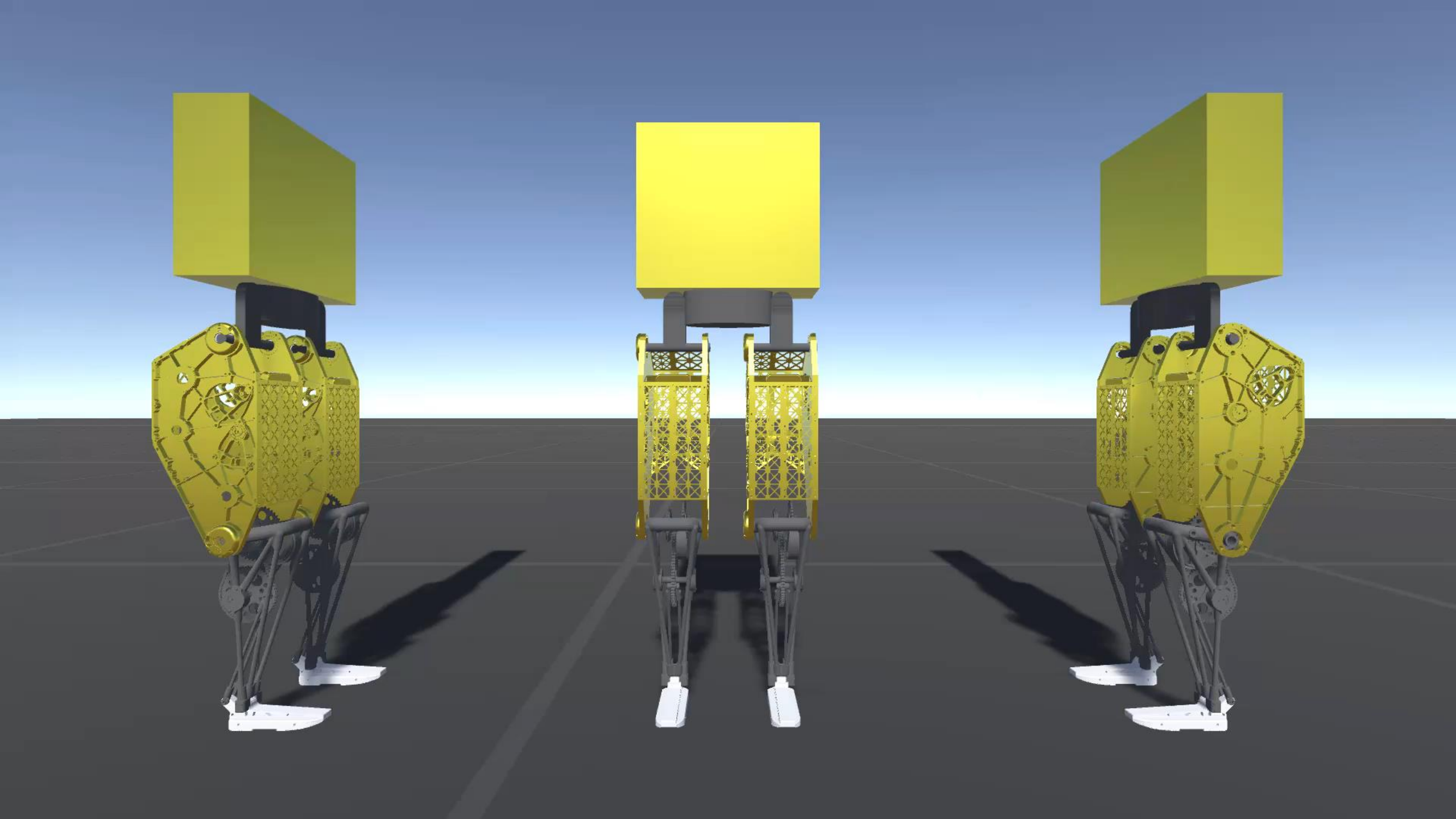


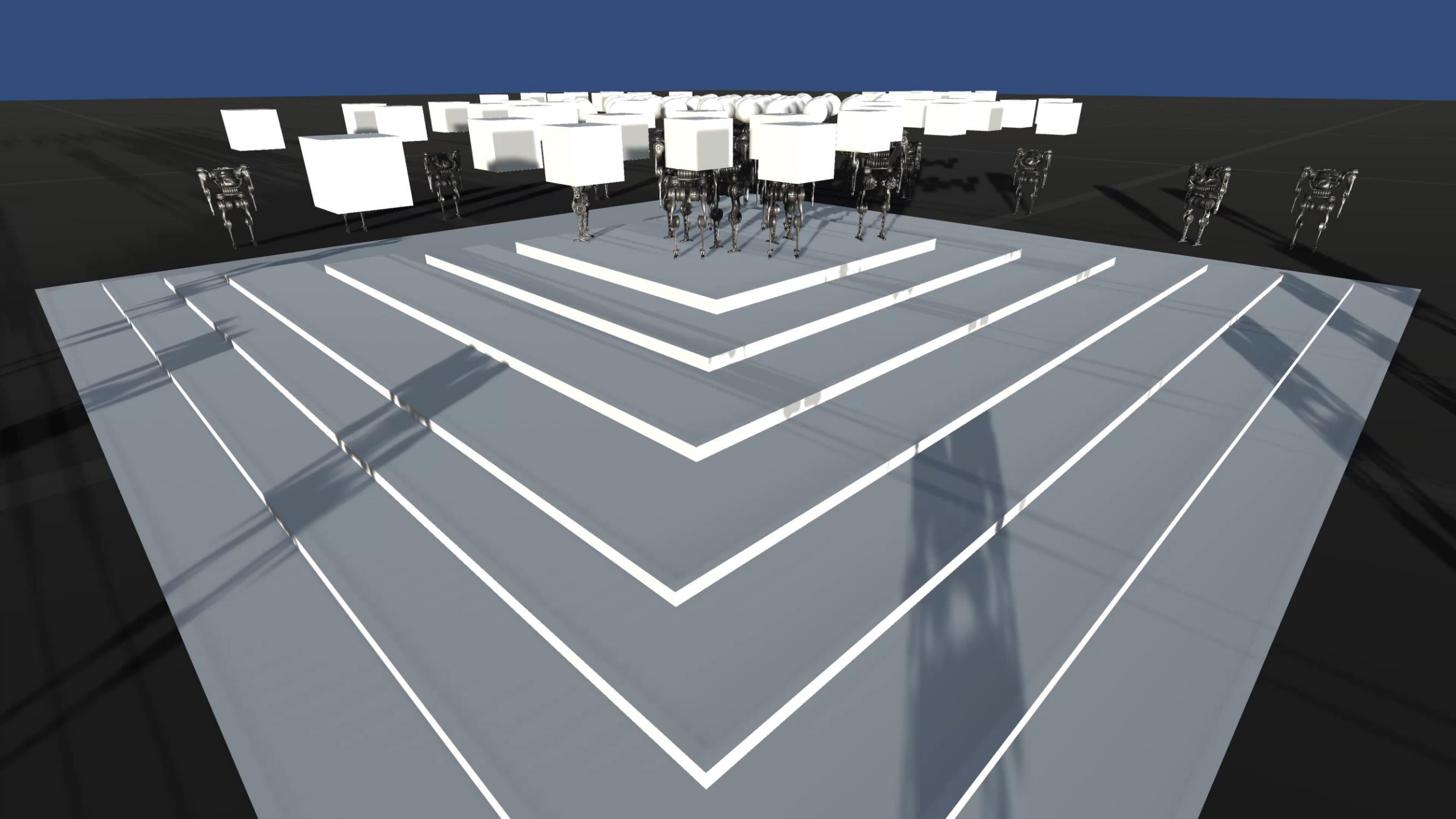





02
Unitree







提纲

THU-SHU ROBOART LAB

一、从人工智能到具身智能的转变

二、知识与数据双驱动的强化学习

三、双足机器人多模运动跟踪控制

四、四足机器人碰撞感知越障控制

AbuDhabi 2024
IROS

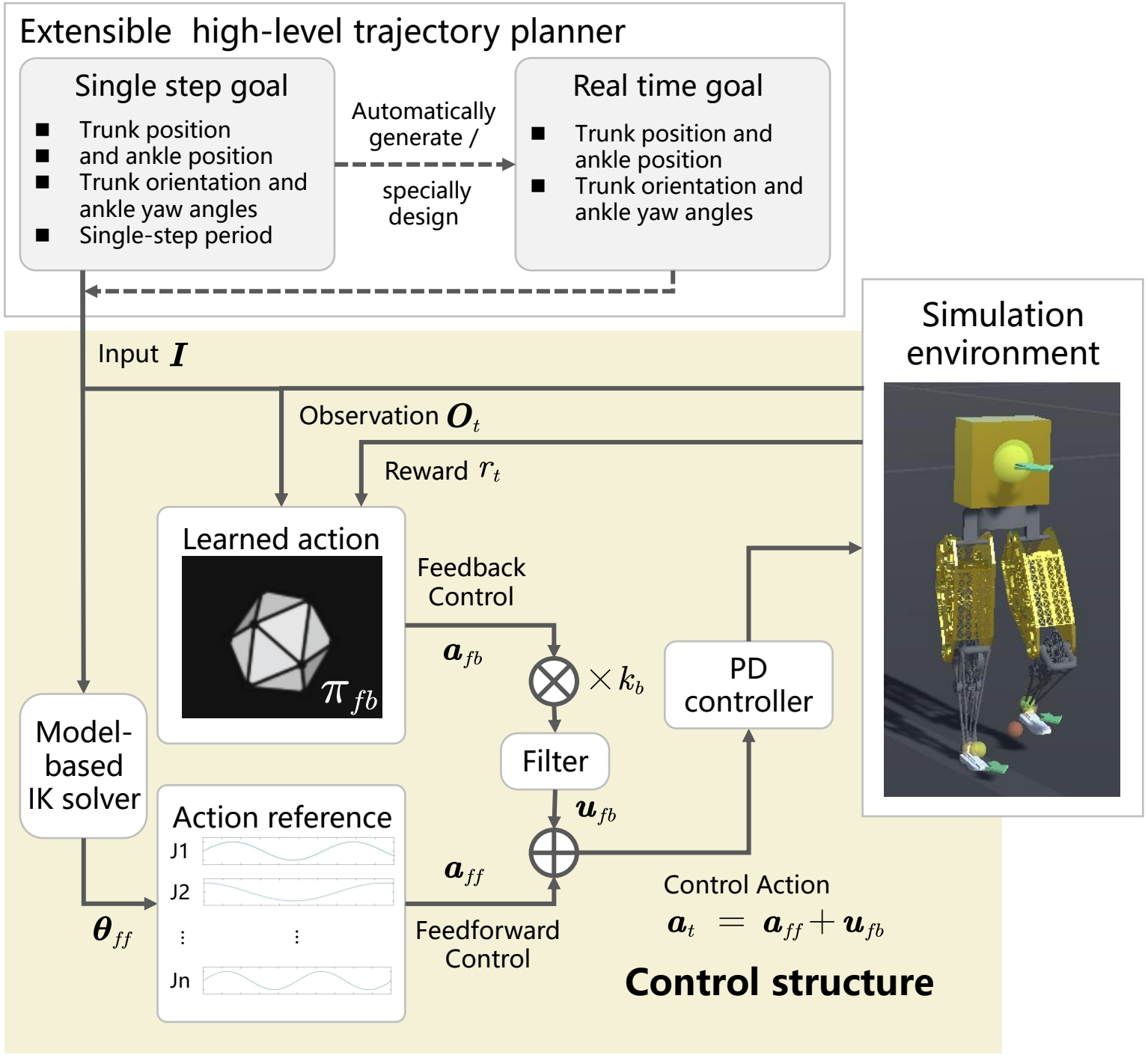
2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 13 – 17, 2024 | Abu Dhabi ,UAE

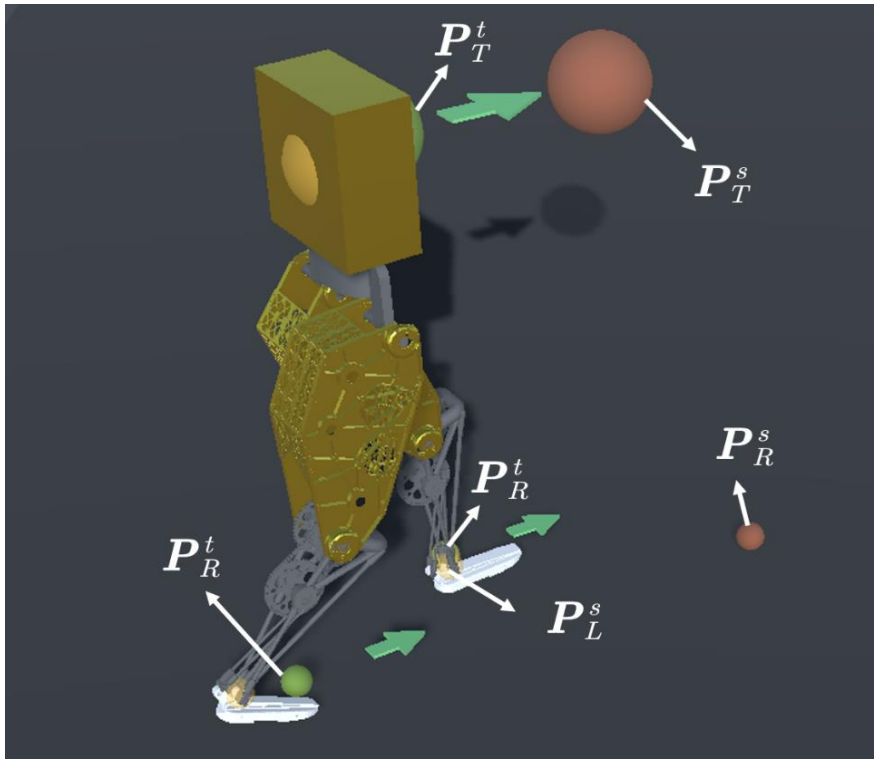
Agile and versatile bipedal robot tracking control through reinforcement learning

Jiayi Li, Linqi Ye, Yi Cheng, Houde Liu*, and Bin Liang
lijiayi21@mails.tsinghua.edu.cn

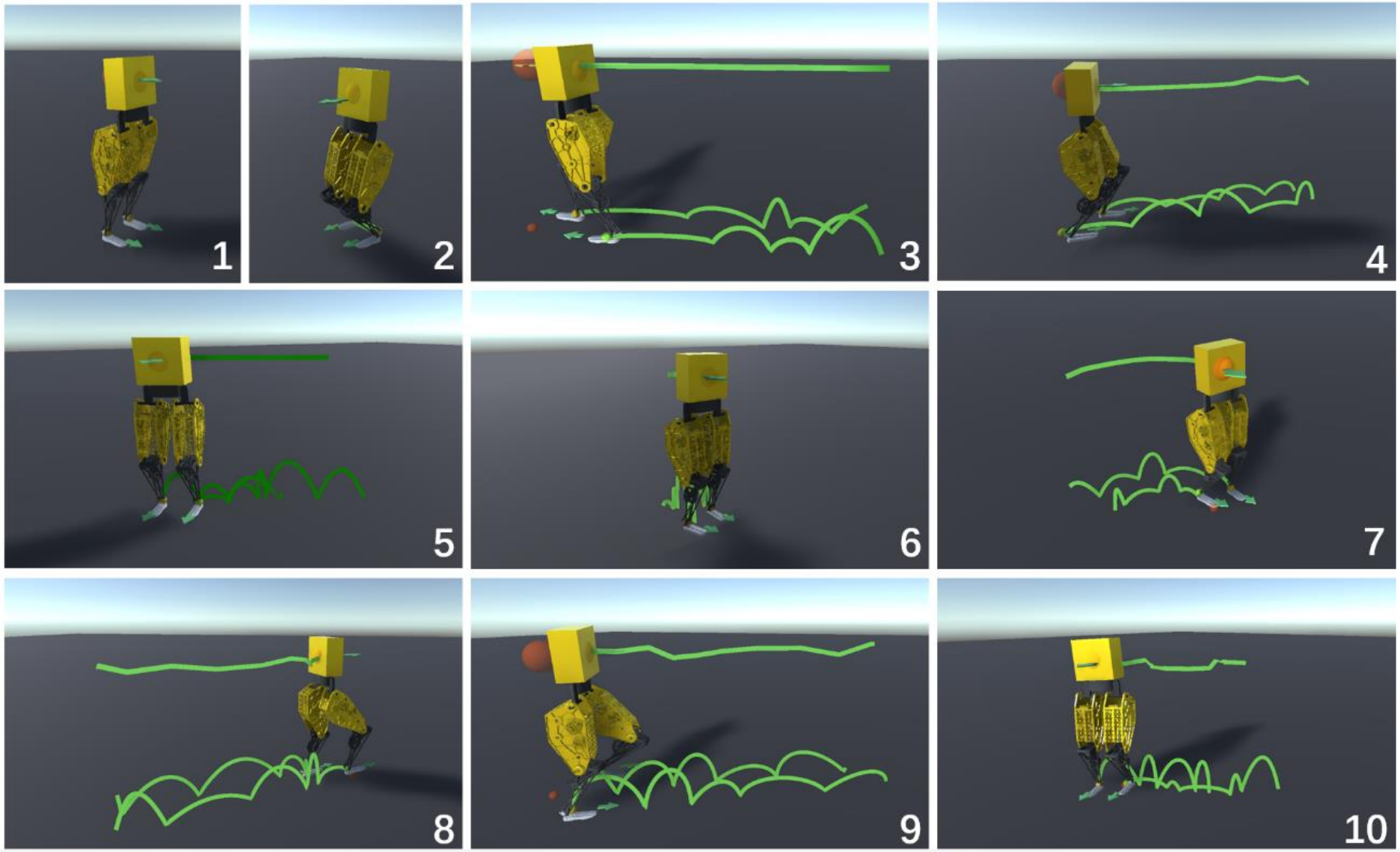
The Center for Artificial Intelligence and Robotics, Tsinghua Shenzhen International Graduate School, Tsinghua University & the Institute of Artificial Intelligence, Collaborative Innovation Center for the Marine Artificial Intelligence, Shanghai University

Control structure

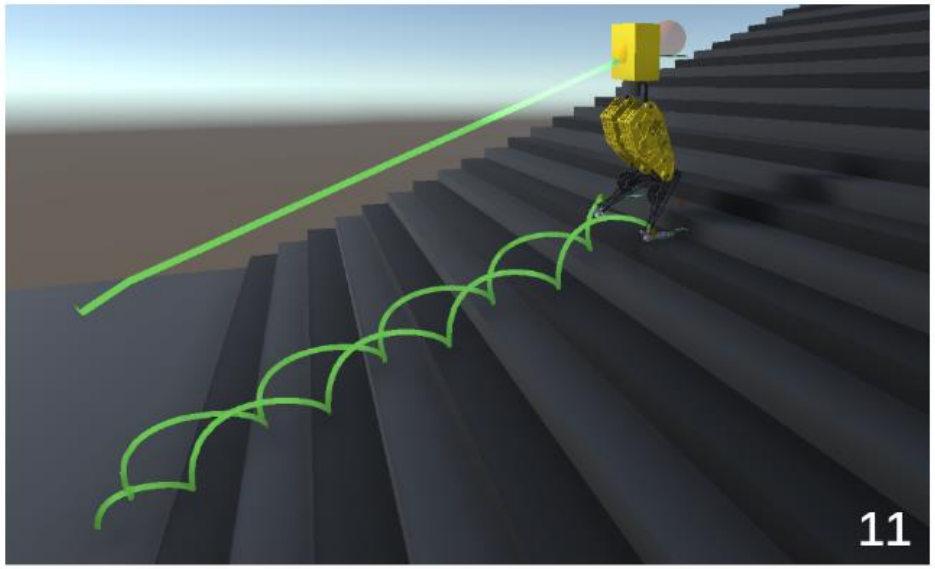


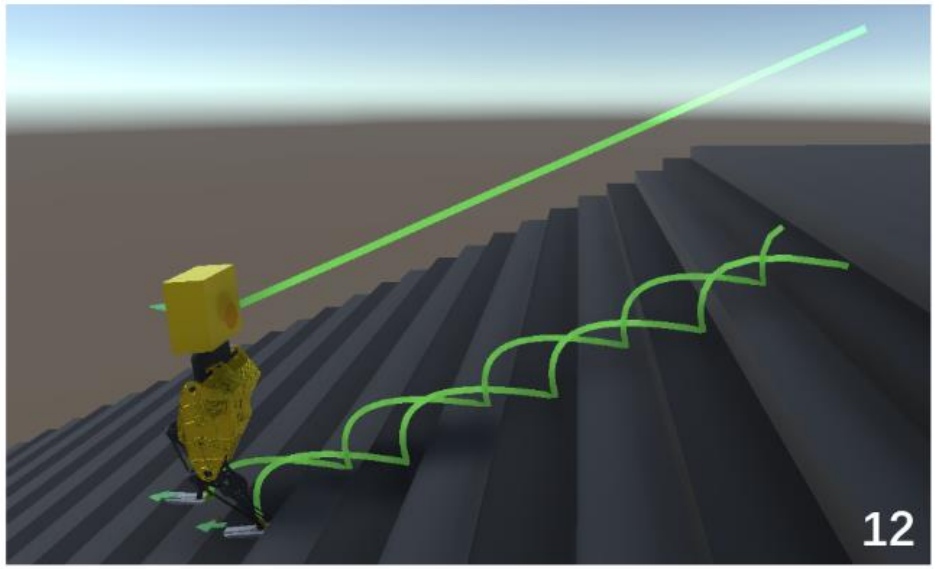


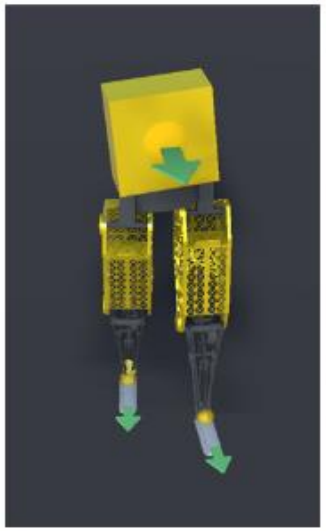
Trajectory planner

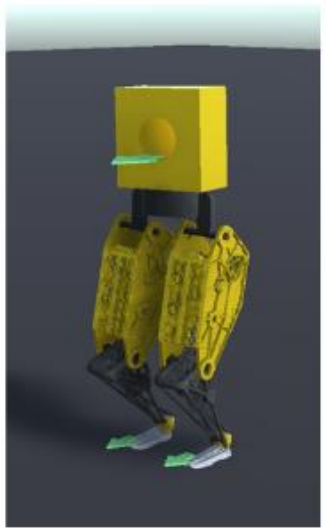


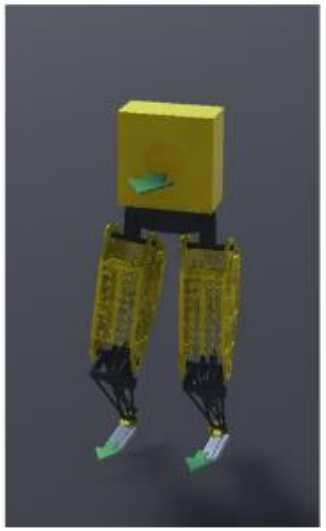
Trajectory planner

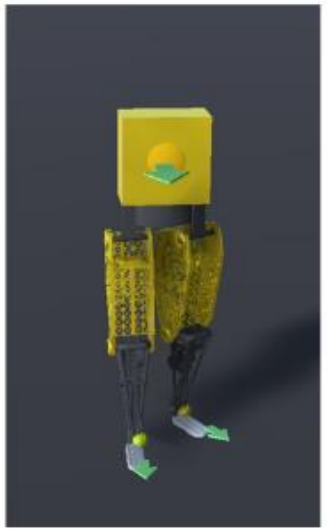


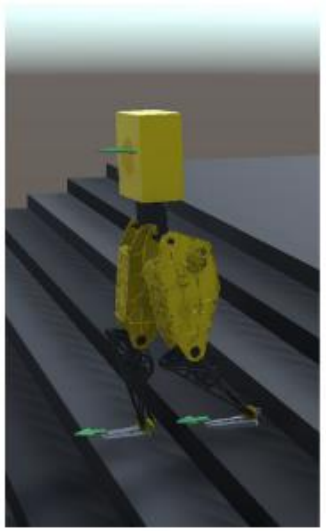




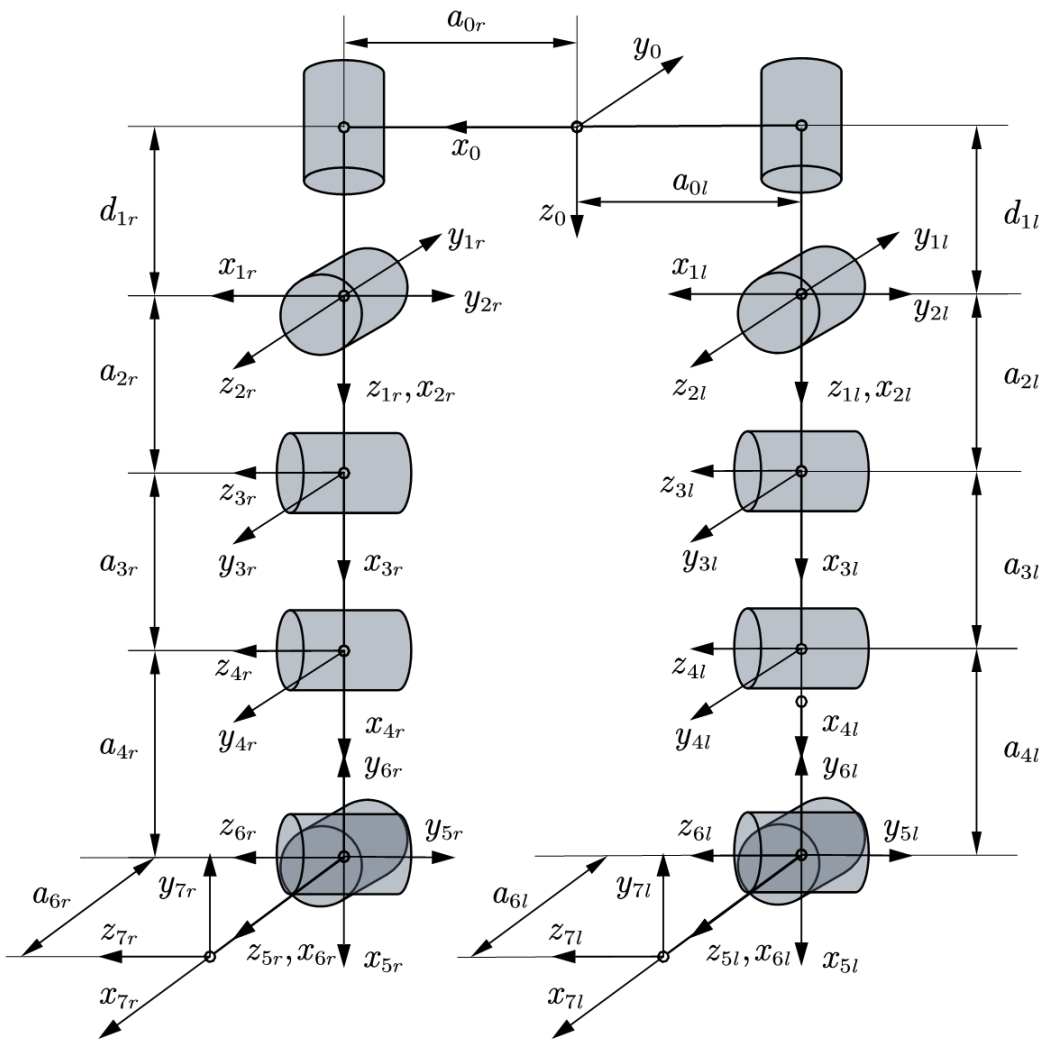








Feedforward signal



$$\boldsymbol{\theta}_{ff} = ikine(\boldsymbol{P}_L^t, \psi_L^t, \boldsymbol{P}_R^t, \psi_R^t)$$
$$\boldsymbol{a}_{ff} = normalization(\boldsymbol{\theta}_{ff})$$

Observation

$$\boldsymbol{O}_t = [\boldsymbol{O}_{input} \quad \boldsymbol{O}_{state} \quad \boldsymbol{a}_{ff}]$$

$$\boldsymbol{O}_{input} = [\boldsymbol{I}^s_{error} \quad \boldsymbol{I}^t_{error} \quad \boldsymbol{I}^t_v \quad \boldsymbol{T}]$$

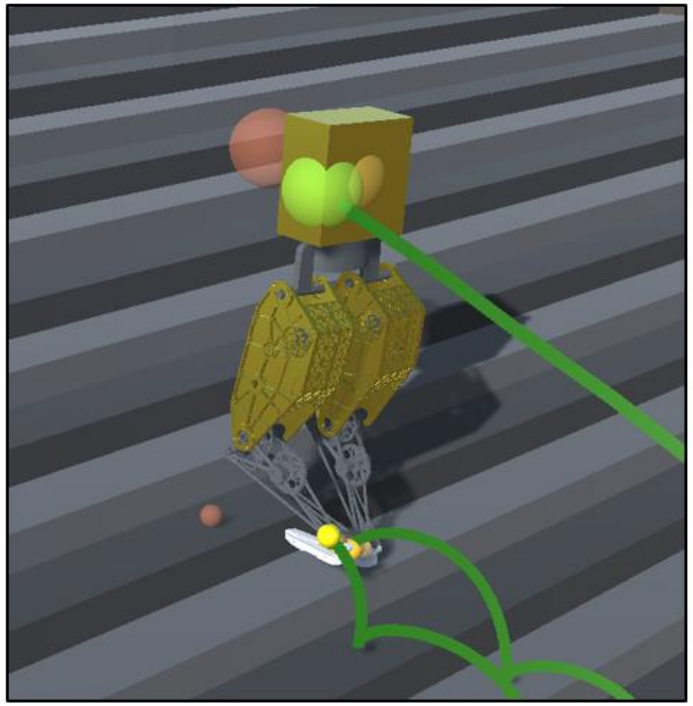
$$\boldsymbol{O}_{state} = [\boldsymbol{g}_T \ \ \boldsymbol{v}_T \ \ \boldsymbol{w}_T \ \ \boldsymbol{J}_P \ \ \boldsymbol{J}_V]$$

Reward

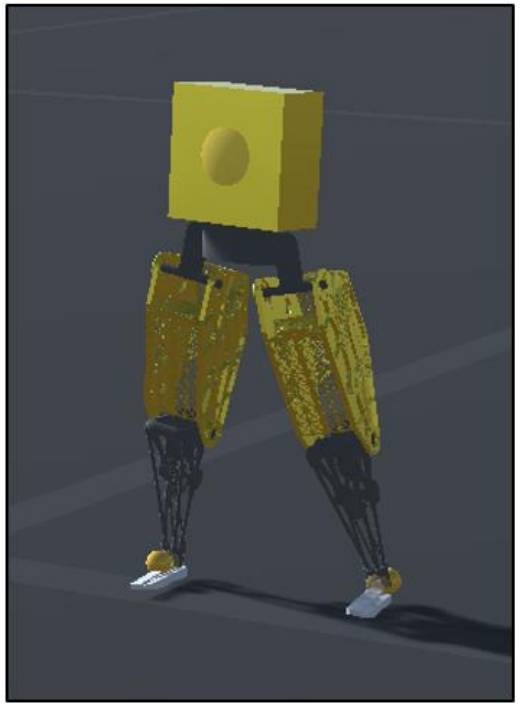
$$r = r_{live} + r^T_{pose} + r^L_{pose} + r^R_{pose} + r^T_{rote} + r^L_{rote} + r^R_{rote}$$
$$+ r_v + r_w + r_{JV} + r_{JA}$$

Curriculum

Level ground task training

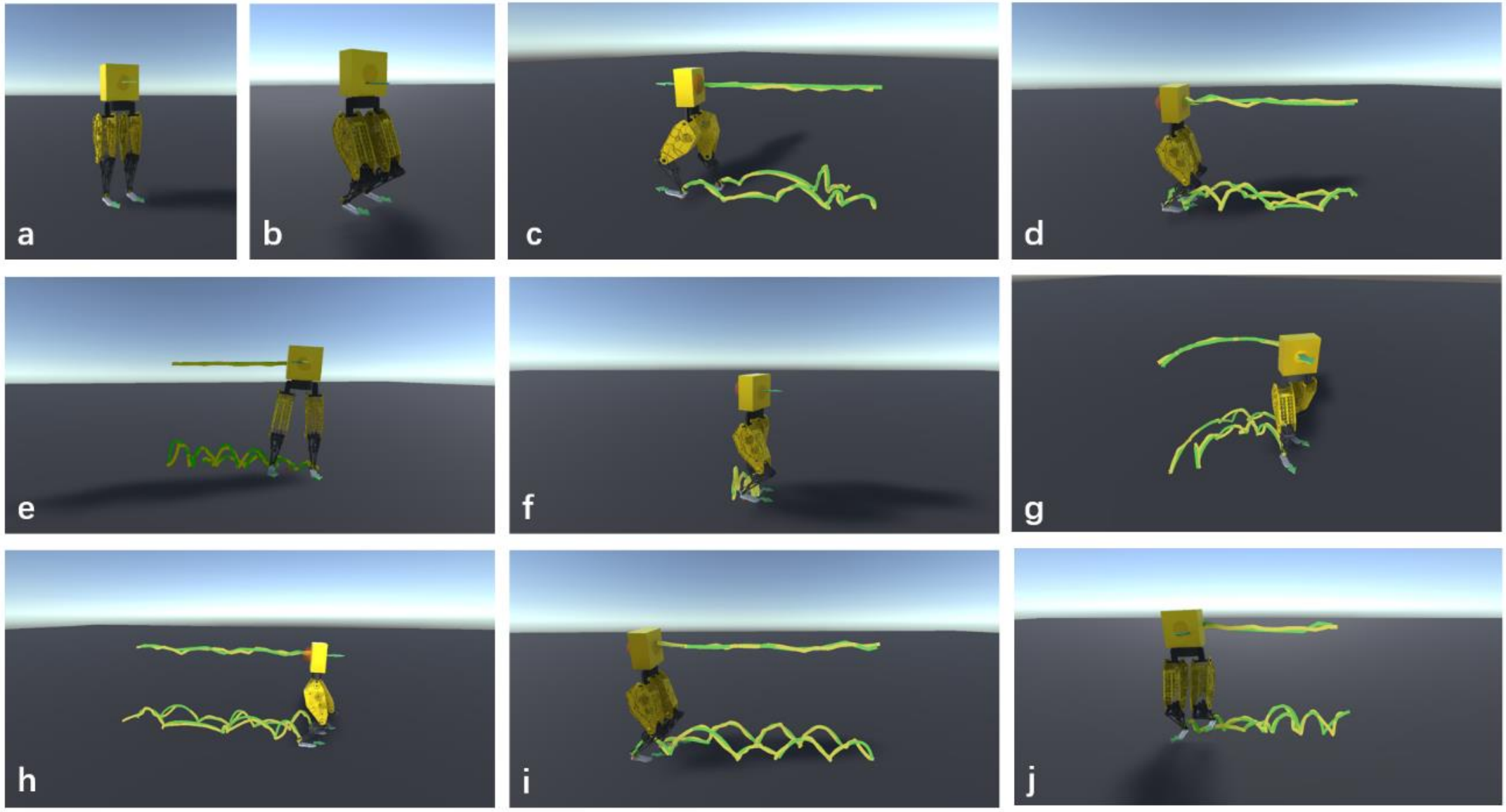


Stair climbing training

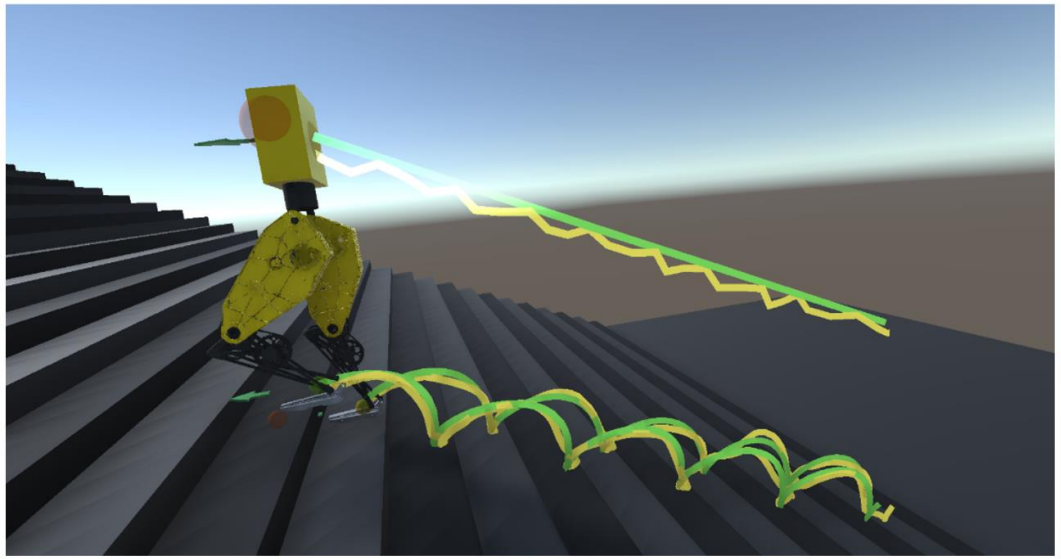


static balance maintenance training

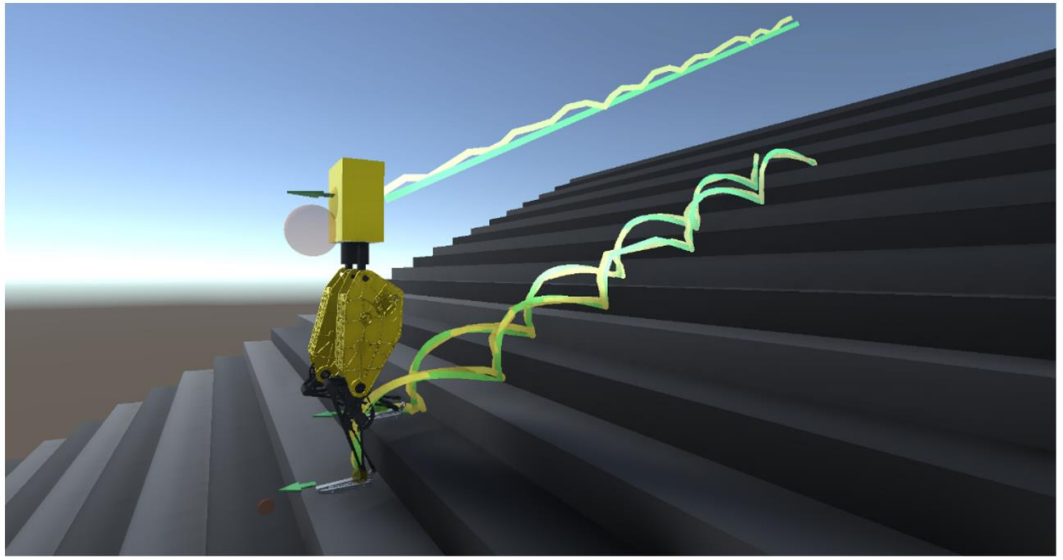
Results



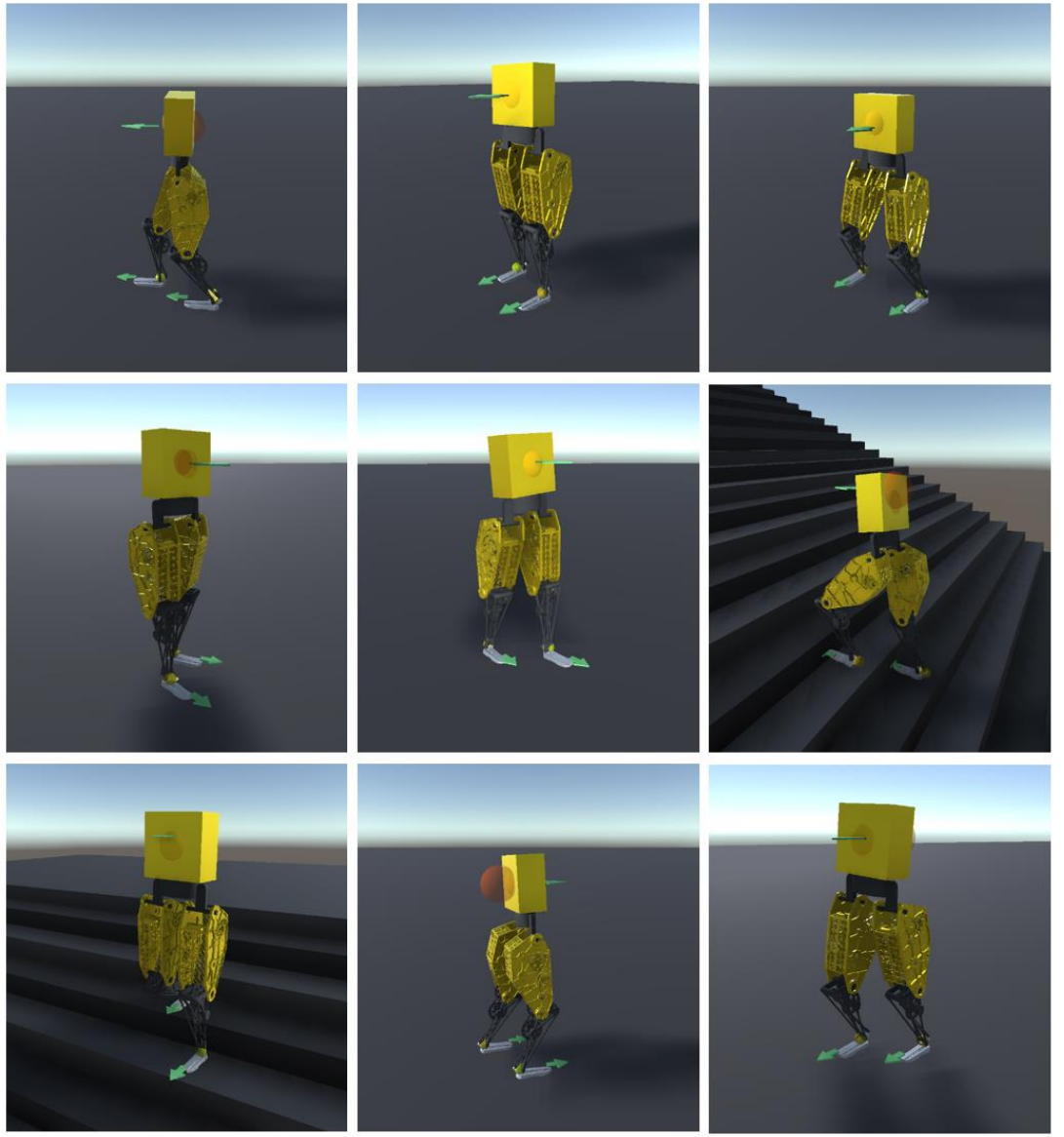
Results



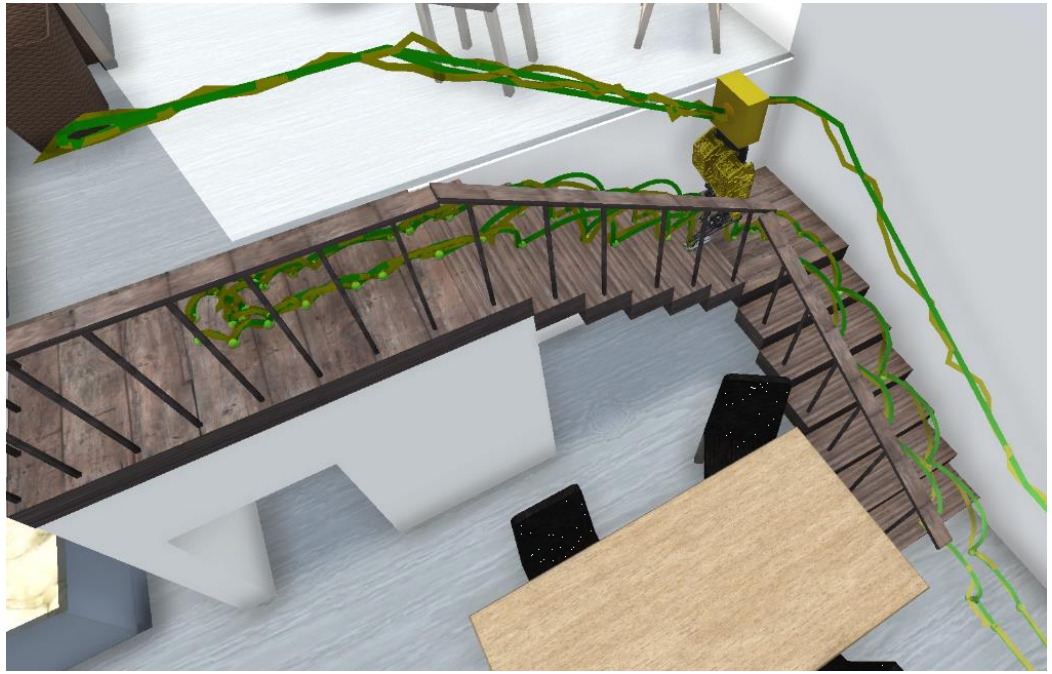
实际轨迹 目标轨迹

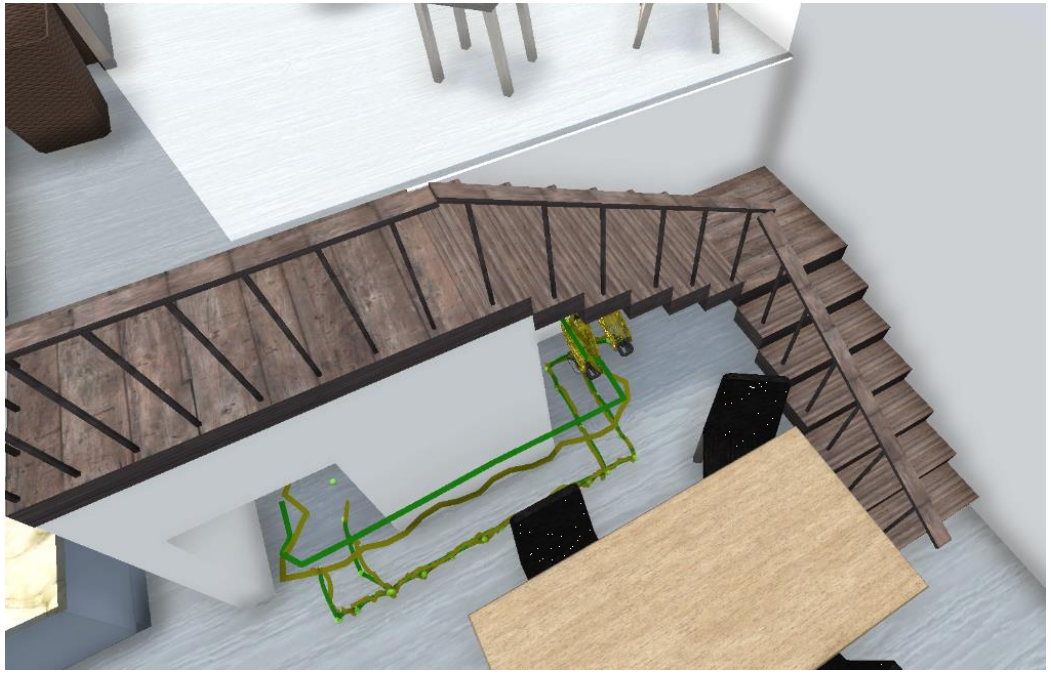


实际轨迹 目标轨迹



Results





提纲

THU-SHU ROBOART LAB

一、从人工智能到具身智能的转变

二、知识与数据双驱动的强化学习

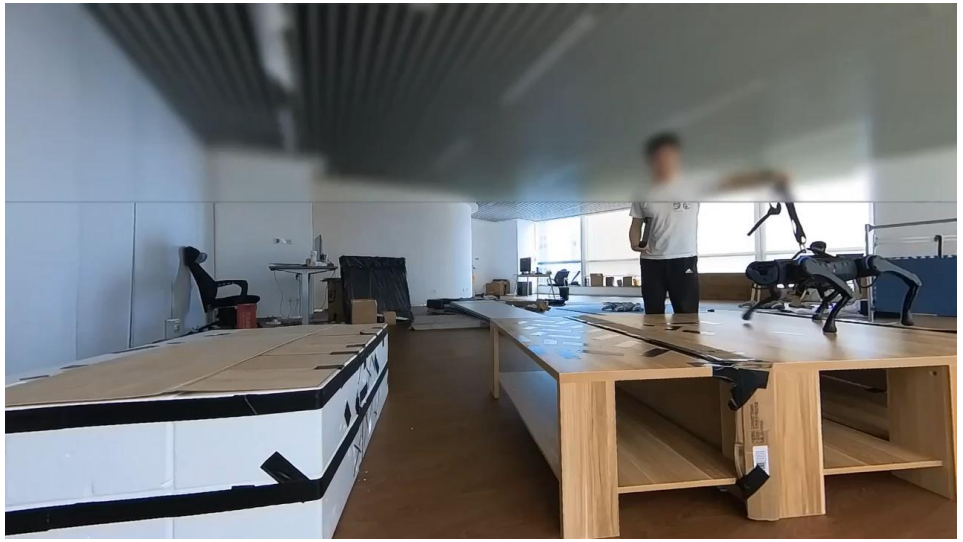
三、双足机器人多模运动跟踪控制

四、四足机器人碰撞感知越障控制

Quadruped robot traversing 3D complex environments with limited perception

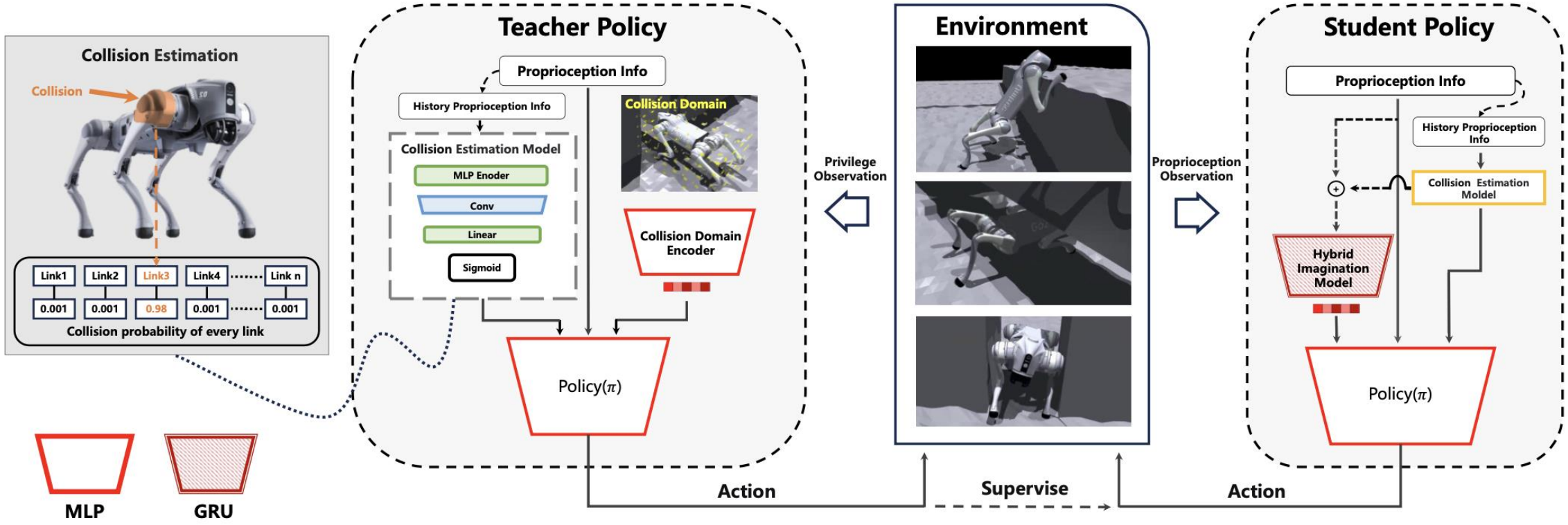
Yi Cheng*, Hang Liu*, Guoping Pan, Linqi Ye, Houde Liu, Bin Liang

Background



It is hard to traversing 3D complex environments without exral perception

Method



Teacher-student based two-stage end-to-end training framework

Method

Observation:

$$s_t = \begin{bmatrix} o_t & \hat{c}_t & \hat{v}_t & p_t & e_t \end{bmatrix}^{\mathrm{T}}$$

Pos reward:

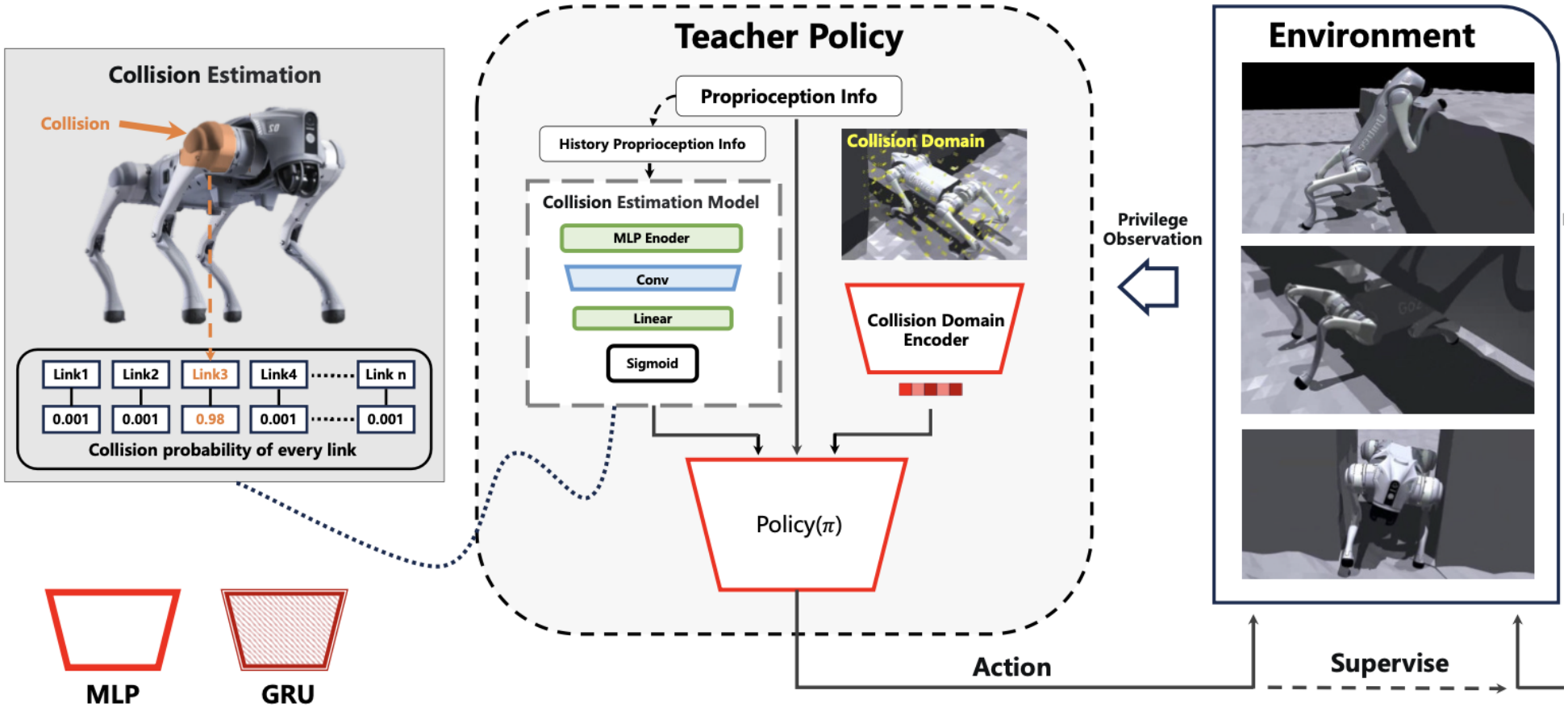
$$\begin{aligned} r_{Pos} &= W_1 * r_{GuidePos} + W_2 * r_{NaturalPos} \\ &= W_1 * (|q_{hip}^{right} + q_{hip}^{left}|^2 + |q_{hip}^{front} - q_{hip}^{behind}|^2) \\ &+ W_2 * |q_{dof}^{default} - q_{dof}|^2 \end{aligned}$$

Vel Reward:

$$r_{vel} = L_{d_v} \frac{min(v \cdot cos(\theta_{yaw}^{cmd} - \theta_{yaw}), v^{cmd})}{v^{cmd}}$$

$$L_{d_v} = \begin{cases} 1, & v \cdot cos(\theta_{yaw}^{cmd} - \theta_{yaw}) > 0 \\ -3, & v \cdot cos(\theta_{yaw}^{cmd} - \theta_{yaw}) < 0 \end{cases}$$

Method

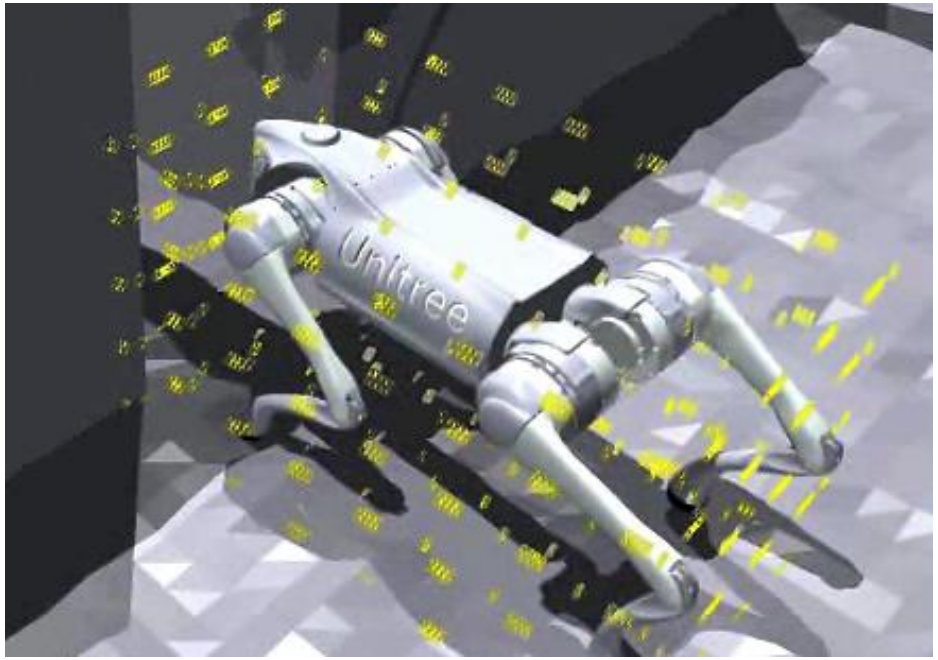


Estimation: $\hat{c}_t = \varphi(x_{t-k}, x_{t-k+1} : x_t)$

Loss: $BCE(\hat{c}_t, c_t) = -(c_t log\hat{c}_t + (1 - c_t)log(1 - \hat{c}_t))$

Method

Collision Domain

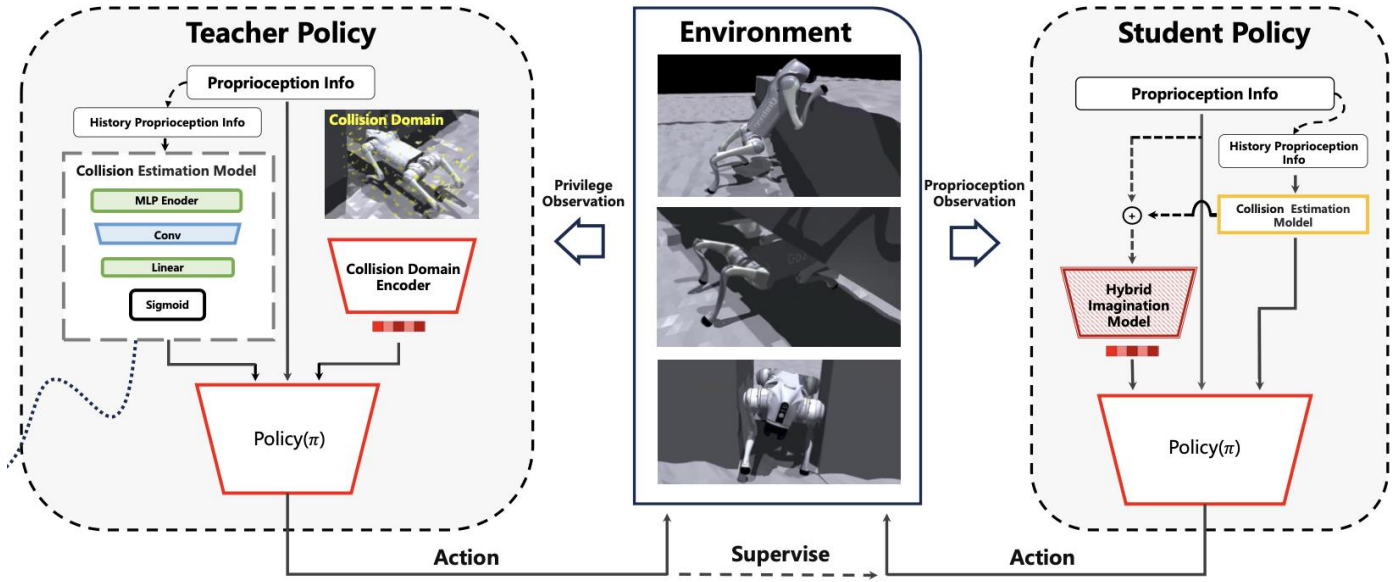


Properties	Obstacle Train Ranges (m) ($[l_{easy}, l_{hard}]$)	Test Ranges (m) ($[l_{easy}, l_{hard}]$)
Highland	[0.05, 0.55]	[0.25, 0.55]
Barrier	[0.31, 0.00]	[0.16, 0.00]
Tunnel	[0.40, 0.25]	[0.38, 0.25]
Crack	[0.38, 0.28]	[0.32, 0.28]

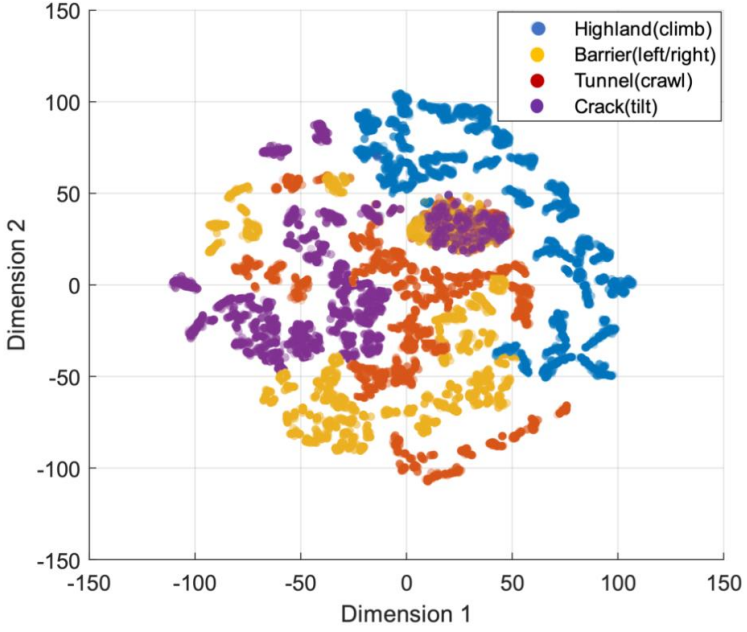
Properties	Parameters Go2 Body (m)	Collision Domain (m)
length	0.71	0.90
width	0.32	0.40
height	0.40	0.50

Method

Implicit Collision Domain imagination



T-SNE Analysis

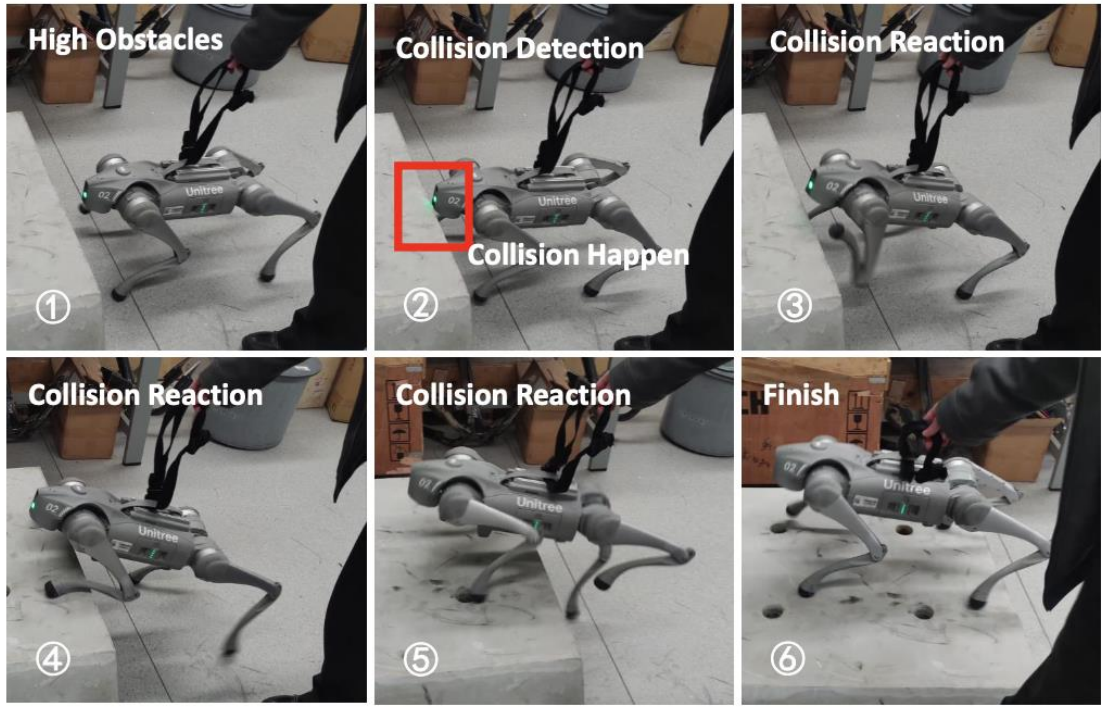


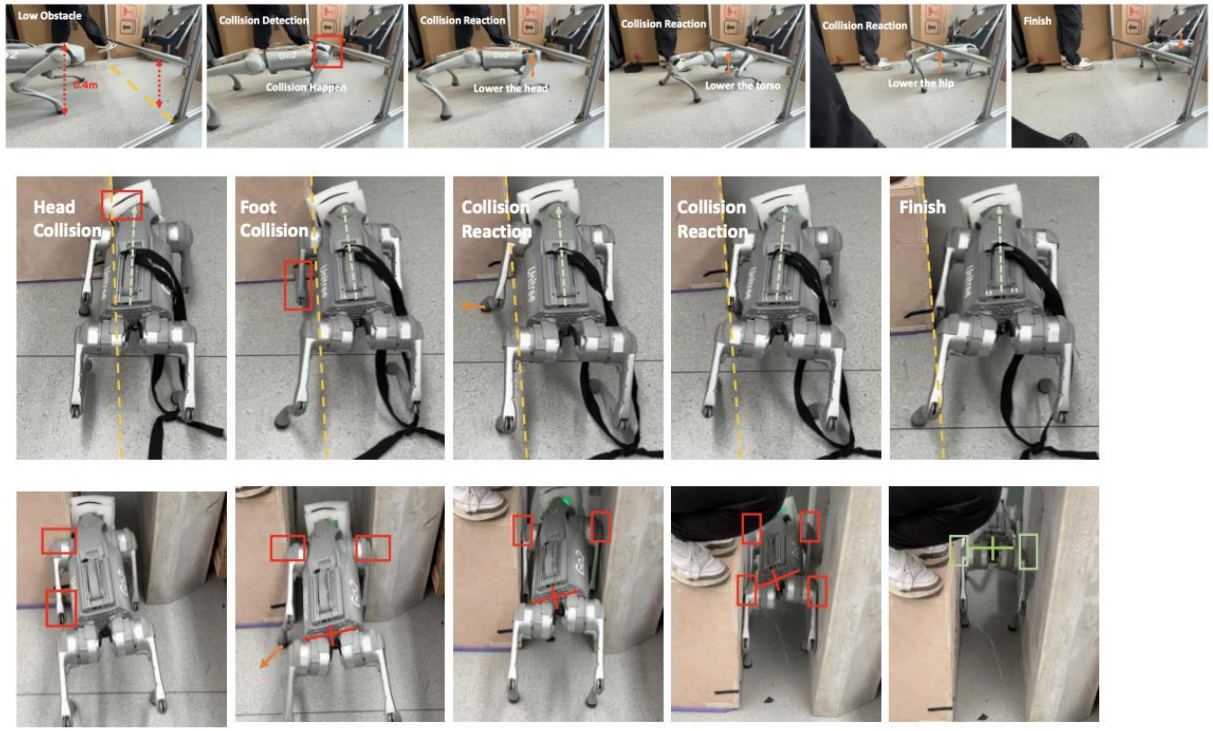
Experiment

	Success Rate ↑				Average Displacement↑			
	Highland	Barrier	Tunnel	Crack	Highland	Barrier	Tunnel	Crack
Baseline	0.00	0.12	0.00	0.00	0.004	0.132	0.005	0.006
RMA	0.27	0.61	0.67	0.53	0.256	0.550	0.614	0.497
WTW	0.00	0.11	1.00	0.00	0.004	0.103	1.000	0.005
Ours w/o R.V	0.00	0.31	0.30	0.00	0.003	0.197	0.218	0.005
Ours w/o Col	0.89	0.81	0.89	0.76	0.769	0.771	0.898	0.625
Ours w/o H.O	0.93	0.87	0.94	0.83	0.796	0.793	0.959	0.711
Ours	0.94	0.92	0.96	0.89	0.821	0.853	0.979	0.798
Teacher	0.99	1.00	1.00	1.00	1.000	1.000	1.000	1.000

- Ours w/o R.V: We use the ordinary linear velocity tracking reward [25] instead of the linear velocity track- ing reward with heading constraints.
- Ours w/o Col: Training without collision estimator.
- Ours w/o H.O: Training collision estimator without history observation.
- Baseline: Training directly with only proprioception.
- RMA[21]: Employing an Adaptation Module to Estimate All Privileged Observations Including Terrain
- WTW(Walk-These-Ways)[23]: Leverages expert knowledge realize Multiplicity of Behavior.

Experiment


High Obstacles
①
Collision Detection
Collision Happen
②
Collision Reaction
③
Collision Reaction
④
Collision Reaction
⑤
Finish
⑥


Low Obstacle
Collision Detection
Collision Happen
Collision Reaction
Lower the head
Collision Reaction
Lower the torso
Collision Reaction
Lower the hip
Finish
Head Collision
Foot Collision
Collision Reaction
Collision Reaction
Finish


Highland
Height (m)
Go2 center height
Success rate
1
0.5
0
0.1
0.2
0.3
0.4
Barrier
Width (m)
Go2 half of width
0
0.04
0.08
0.12
0.16
Tunnel
Height (m)
Go2 top height
0.4
0.35
0.32
0.3
0.27
Crack
Width (m)
Go2 width
0.35
0.32
0.3
0.28
Ours
Go2-default
Go2-special
WTW

Quadruped robot traversing 3D complex environments with limited perception

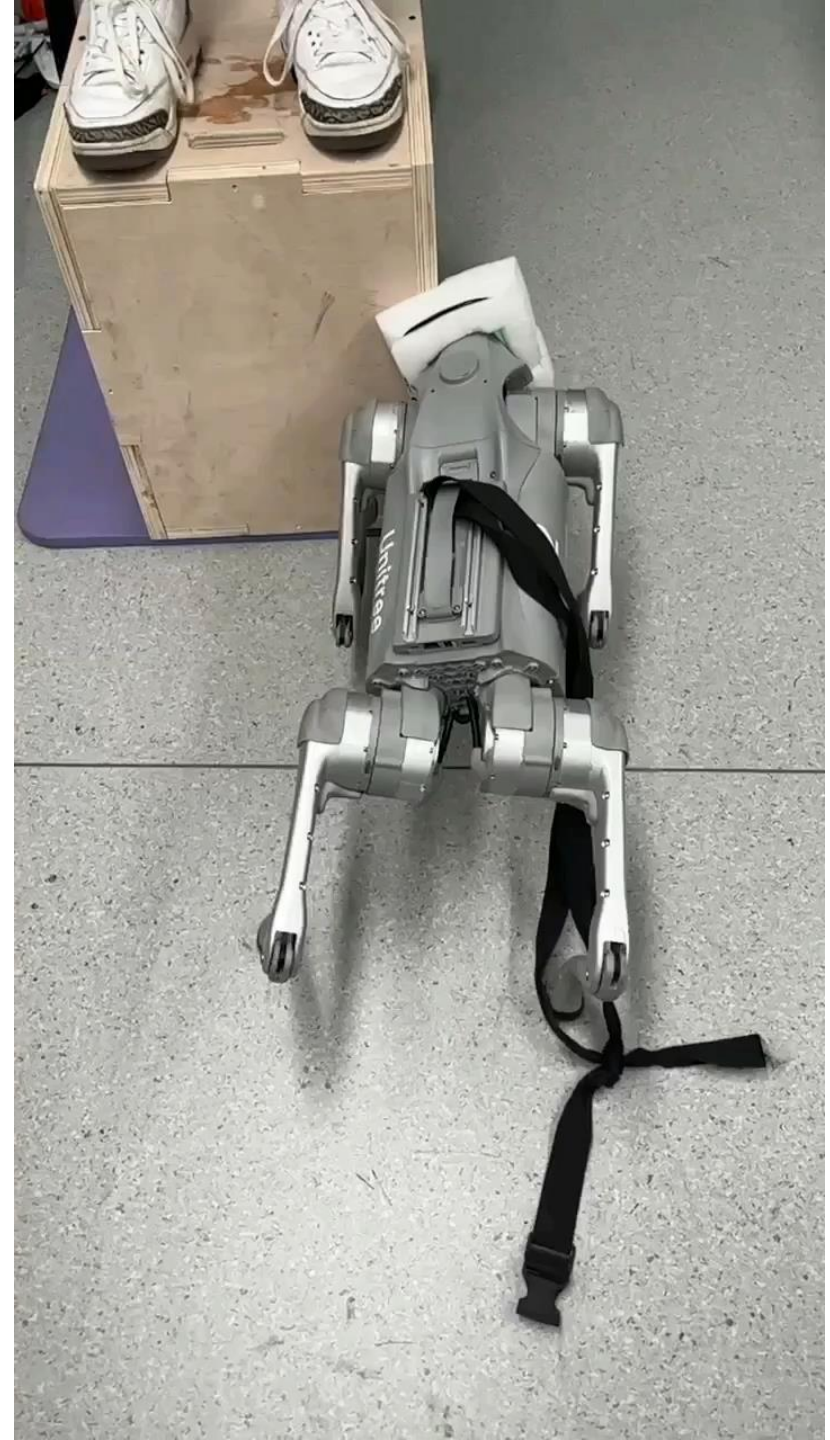
Yi Cheng*, Hang Liu*, Guoping Pan, Linqi Ye, Houde Liu, Bin Liang

AbuDhabi 2024
IROS


Go2


ZSCH
02
Unitree


Unitree

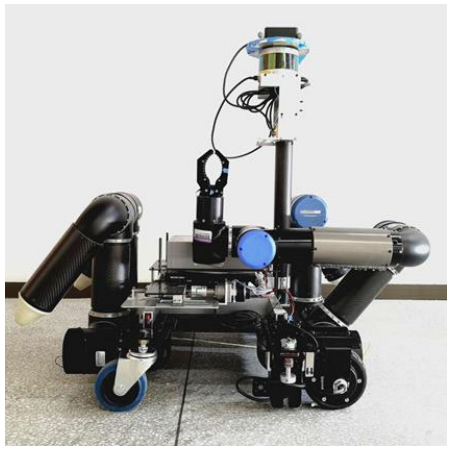

Unitree

AI&ROBOT
清华大学深圳国际研究生院 智能机器人实验室

上海大学人工智能研究院
SCHOOL OF ARTIFICIAL INTELLIGENCE SHANGHAI UNIVERSITY

上海大学未来技术学院
SCHOOL OF FUTURE TECHNOLOGY, SHANGHAI UNIVERSITY

智能机器人实验室 || 具身智能研究团队


“清华-上大”机器艺术与具身智能实验室，聚焦机器人、具身智能及与艺术结合的前沿研究。

（网址： https://linqi-ye.github.io/ ）

THU-SHU ROBOART LAB

谢谢大家