Bibliography Library for Quadruped Learning

Xinyu Zhang, Zhiyuan Xiao

2023.09 - 2023.11.16

关键词

添加模块:
【MPC】【KMP】【CPG】【WBC】【步态规划器】【组合策略】【状态估计器】【视觉感知】

使用技巧:
【策略迁移】【域随机化】【模型辨识】【课程学习】【地形课程】【速度课程】【模仿学习】【逆强化学习】【对抗学习】【privileged learning】【teacher-student learning】【元学习】【进化策略】【分层网络】【分层框架】【分段训练】【同时训练】【交替训练】【真实环境训练】【分布式学习】【表征学习】

针对场景:
【崎岖路面】【复杂地形】【机器人机械臂协作】【环境互动】【低重力】

输出动作:
【输出力矩】【输出足端/脚趾轨迹】

达到目的:
【抗噪声】【安全约束】【感知融合】【数据高效】【鲁棒运动】【跳跃运动】【节能运动】【步态转换】【简单奖励】

课题组:
【Hutter】ETH 【Levine】UCB 【Jie Tan】Google 【Ijspeert】EPFL

Science Robotics – Marco Hutter

[2019] Hwangbo 等人聚焦仅利用本体传感器信息学习四足机器人的运动控制，提出利用真实数据训练电机模型并将其并应用于仿真环境中以提高物理仿真环境的保真度。同时采用MLP 作为策略网络，通过堆叠历史关节信息推断足端触地信息。[2020] Lee 等人改用PMTG 架构以及采用特权学习，并将训练过程分为两个阶段：第一阶段训练教师网络。该网络由两个MLP 网络构成，其中一个MLP 网络为策略网络，另一个网络充当编码器，负责将机器人不易获取的（地形高程图、地面摩擦系数、足端状态等）信息进行编码之后作为策略网络的输入；第二阶段采用模仿学习训练学生网络。学生网络同样由策略网络和编码器两个网络构成，前者直接复制教师网络的策略网络，后者改用序列模型TCN，并以历史传感器信息作为该网络的输入。特权学习的实质是知识蒸馏。[2022] Miki 等人考虑结合外部传感器信息的四足机器人鲁棒运动控制，提出了基于Attention 机制的belief state encoder，可以融合感知信息和本体信息，并在感知模块失效时通过历史本体信息修正对地形的估计。

[2019] Learning agile and dynamic motor skills for legged robots
[2020] Learning quadrupedal locomotion over challenging terrain
[2022] Learning robust perceptive locomotion for quadrupedal robots in the wild
[2023] Scientific exploration of challenging planetary analog environments with a team of legged robots*

Science Robotics – Jemin Hwangbo

[2023] Learning quadrupedal locomotion on deformable terrain

Science Robotics – Zhibin Li

[2020] Multi-expert learning of adaptive legged locomotion

Nature Machine Intelligence

[2022] High-speed quadrupedal locomotion by imitation-relaxation reinforcement learning
[2023] Identifying important sensory feedback for learning locomotion skills

Nature Scientific Reports

[2022 NSR] Controlling the Solo12 quadruped robot with deep reinforcement learning

Parallel Training Platform

[2022 CoRL] Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning

Privileged Learning: Teacher-Student Policy

[2020 SCI RO] Learning quadrupedal locomotion over challenging terrain
[2021 IROS] Terrain-Aware Risk-Assessment-Network-Aided Deep Reinforcement Learning for Quadrupedal Locomotion in Tough Terrain
[2022 SCI RO] Learning robust perceptive locomotion for quadrupedal robots in the wild
[2022 RSS] Rapid Locomotion via Reinforcement Learning
[2022 RAL] Reinforcement Learning With Evolutionary Trajectory Generator: A General Approach for Quadrupedal Locomotion
[2023 SCI RO] Scientific exploration of challenging planetary analog environments with a team of legged robots
[2023 RAL] Learning Robust and Agile Legged Locomotion Using Adversarial Motion Priors
[2023 RAL] Learning-Based Design and Control for Quadrupedal Robots With Parallel-Elastic Actuators
[2023 CoRL] Legged Locomotion in Challenging Terrains using Egocentric Vision
[2023 arXiv] Extreme Parkour with Legged Robots
[2023 arXiv] Learning Vision-based Pursuit-Evasion Robot Policies

RMA Architecture

快速适应模块
Kumar 等人提出RMA 框架，该框架分为两个训练阶段。第一阶段训练可以接触（编码后的）环境信息、动力学参数信息的教师策略，教师网络由两个MLP 构成，其中一个MLP 作为策略网络，另一个MLP 充当编码器，编码后的信息将作为输入策略网络中。第二阶段训练学生网络，先固定策略MLP 网络，改用以历史状态信息作为输入的TCN 网络作为环境适应模块并代替教师网络中的编码器，采用监督学习训练TCN[1]。后续拓展工作还引进了第三阶段的训练，以及提出同时训练教师网络和学生网络的方法。[2, 3]

容错运动策略
Luo 等人提出了故障恢复和故障容错控制框架FT-Net，该框架采用RMA 架构，将参数化的不同故障场景进行隐式编码，并通过adaptor 统一调节故障情况下的运动控制策略。[4]

深度相机
Kareer 等人提出基于视觉的ViNL 方法。不同于Kumar 采用历史状态信息，Kareer 采用深度相机信息进行知识蒸馏[5]。

[2021 RSS] RMA: Rapid Motor Adaptation for Legged Robots
[2021 CoRL] Minimizing Energy Consumption Leads to the Emergence of Gaits in Legged Robots
[2022 IROS] Adapting Rapid Motor Adaptation for Bipedal Robots
[2023 CoRL] Legged Locomotion in Challenging Terrains using Egocentric Vision
[2023 CoRL] Deep Whole-Body Control: Learning a Unified Policy for Manipulation and Locomotion
[2023 RAL] FT-Net: Learning Failure Recovery and Fault-tolerant Locomotion for Quadruped Robots
[2023 ICRA] Legs as Manipulator: Pushing Quadrupedal Agility Beyond Locomotion
[2023 ICRA] DreamWaQ: Learning Robust Quadrupedal Locomotion With Implicit Terrain Imagination via Deep Reinforcement Learning
[2023 ICRA] ViNL: Visual Navigation and Locomotion Over Obstacles

PMTG Architecture

[2018 CoRL] Policies modulating trajectory generators
[2019 IROS] Hierarchical Reinforcement Learning for Quadruped Locomotion
[2020 SCI RO] Learning quadrupedal locomotion over challenging terrain
[2020 CoRL] Data Efficient Reinforcement Learning for Legged Robots
[2021 CoRL] From Pixels to Legs: Hierarchical Learning of Quadruped Locomotion
[2021 ICRA] Real-Time Trajectory Adaptation for Quadrupedal Locomotion using Deep Reinforcement Learning
[2021 IROS] Terrain-Aware Risk-Assessment-Network-Aided Deep Reinforcement Learning for Quadrupedal Locomotion in Tough Terrain
[2022 SCI RO] Learning robust perceptive locomotion for quadrupedal robots in the wild
[2022 RAL] CPG-RL: Learning Central Pattern Generators for Quadruped Locomotion
[2022 RAL] Reinforcement Learning With Evolutionary Trajectory Generator: A General Approach for Quadrupedal Locomotion
[2022 arXiv] Visual CPG-RL: Learning Central Pattern Generators for Visually-Guided Quadruped Navigation
[2022 CoRL] Visual-Locomotion: Learning to Walk on Complex Terrains with Vision
[2023 ICRA] Efficient Learning of Locomotion Skills through the Discovery of Diverse Environmental Trajectory Generator Priors

CPG-RL

[2021 TNNLS] Generic Neural Locomotion Control Framework for Legged Robots
[2021 RAL] CPG-Based Hierarchical Locomotion Control for Modular Quadrupedal Robots Using Deep Reinforcement Learning
[2022 RAL] CPG-RL: Learning Central Pattern Generators for Quadruped Locomotion
[2022 RAL] Reinforcement Learning With Evolutionary Trajectory Generator: A General Approach for Quadrupedal Locomotion
[2022 RAL] Learning Free Gait Transition for Quadruped Robots Via Phase-Guided Controller
[2022 arXiv] Visual CPG-RL: Learning Central Pattern Generators for Visually-Guided Quadruped Navigation

[2023 ICRA] Puppeteer and Marionette: Learning Anticipatory Quadrupedal Locomotion Based on Interactions of a Central Pattern Generator and Supraspinal Drive
[2023 arXiv] Learning Quadruped Locomotion using Bio-Inspired Neural Networks with Intrinsic Rhythmicity

Model-based & Model-free

模仿学习

Carius 等人提出通过最小化control Hamiltonian 的策略搜索方法MPC-Net，可以将模型预测控制器克隆成神经网络策略。[6]。Reske等人使用模仿学习方法，扩展了MPC-net，使用混合专家网络mixture-of-experts network（MEN）实现在单一策略下学习多种步态[7]。Xie 等人使用模型复杂度较低的中心动力学模型来训练强化学习模型[8]。Anne 等人利用MPC 选择使预测奖励最大化的动作序列，并且让智能体采取第一个动作[9]。Shirwatkar 等人提出基于质心动力学的MPC 方法用于生成参考轨迹数据，然后使用模仿学习训练线性策略，以最小化与参考轨迹的偏差。[10]。Kang 等人提出采用Motion imitation 方法，其奖励函数量化了机器人的状态与基于model-based optimal control (MOC) 的运动生成器生成的参考运动之间的相似性[11]。

分层框架

在Yao 等人提出的分层框架中，最优控制作为下层的控制器计算支撑腿各关节的力矩[12]。Yin 等人提出了基于学习的Whole-body Control（WBC）框架，其中上层的神经网络规划期望步态和足端位置（foothold），而基于多刚体动力学的底层控制器输出每个关节的期望力矩[13]。在Margolis 等人提出了基于深度的脉冲控制（DIC）方法中，策略网络生成动作输入到Whole-body Trajectory Generator中生成机身轨迹，再通过底层的MPC+WBIC 控制器追踪生成的动态机身轨迹完成运动[14]。Gangapurwala 采用trajectory optimization solver (TOWR) 生成参考轨迹，并使用强化学习生成一个轨迹偏差量，最终信号输入到WBC 控制器中[15]。Yang 等人采用分层框架，其中使用强化学习训练高层的步态策略，底层使用模型预测控制器（MPC）输出支撑腿的ground reaction forces (GRF)[16]。Da 等人提出了一种结合基于模型的控制与强化学习的分层框架，该框架由一个上层控制器和一个底层控制器组成，前者指定足端的contact configuration，后者接收contact configuration 作为输入，通过QP 生成GRF[17]。Gangapurwala 使用强化学习策略将本体和外部传感信息和速度指令映射到足端规划中[18]。Pandala 等人提出了一个分层框架，其中上层为输出GRF 的鲁棒模型预测控制（RMPC），并利用深度强化学习来计算神经网络来计算RMPC 未被建模的动力学，底层为基于虚拟约束和QP 的非线性控制器。[19]。Yu 提出基于视觉的分层控制算法，其中上层视觉策略将视觉信号和机器人状态作为输入，并输出机器人所需的立足点和基本运动，底层为基于MPC 的控制器[20]。Xu 等人提出的分层控制框架中，上层网络用于步态选择及输出自适应模型参数，底层算法用于实现鲁棒的步态控制（MPC）[21]。

[2020 CoRL] Data Efficient Reinforcement Learning for Legged Robots
[2020 RAL] MPC-Net: A First Principles Guided Policy Search
[2020 RAL] Guided Constrained Policy Optimization for Dynamic Quadrupedal Robot Locomotion
[2021 arXiv] GLiDE: Generalizable Quadrupedal Locomotion in Diverse Environments with a Centroidal Model
[2021 IROS] Hierarchical Terrain-Aware Control for Quadrupedal Locomotion by Combining Deep Reinforcement Learning and Optimal Control
[2021 IROS] Meta-Learning for Fast Adaptive Locomotion with Uncertainties in Environments and Robot Dynamics
[2021 IROS] Run Like a Dog: Learning Based Whole-Body Control Framework for Quadruped Gait Style Transfer
[2021 IROS] Animal gait using motion matching and model-based control on a quadruped robot
[2021 IROS] Robust Feedback Motion Policy Design Using Reinforcement Learning on a 3D Digit Bipedal Robot
[2021 ICRA] Imitation learning from mpc for quadrupedal multi-gait control
[2021 ICRA] Real-Time Trajectory Adaptation for Quadrupedal Locomotion using Deep Reinforcement Learning
[2021 CoRL] Learning to Jump from Pixels
[2021 CoRL] Fast and Efficient Locomotion via Learned Gait Transitions
[2020 CoRL] Learning a Contact-Adaptive Controller for Robust, Efficient Legged Locomotion
[2022 TRO] RLOC: Terrain-Aware Legged Locomotion Using Reinforcement Learning and Optimal Control
[2022 RAL] Robust Predictive Control for Quadrupedal Locomotion: Learning to Close the Gap between Reduced- and Full-Order Models

[2022 CoRL] Visual-Locomotion: Learning to Walk on Complex Terrains with Vision
[2022 ICRA] Learning Efficient and Robust Multi-Modal Quadruped Locomotion: A Hierarchical Approach
[2022 IROS] Zero-Shot Retargeting of Learned Quadruped Locomotion Policies Using Hybrid Kinodynamic Model Predictive Control
[2023 ICRA] Force control for Robust Quadruped Locomotion: A Linear Policy Approach
[2023 arXiv] RL + Model-based Control: Using On-demand Optimal Control to Learn Versatile Legged Locomotion
[2023 RAL] Imitating and Finetuning Model Predictive Control for Robust and Symmetric Quadrupedal Locomotion

Advanced Skills

Rapid locomotion
[2022 RSS] Rapid Locomotion via Reinforcement Learning
[2022 NMI] High-speed quadrupedal locomotion by imitation-relaxation reinforcement learning
[2022 RAL] Concurrent Training of a Control Policy and a State Estimator for Dynamic and Robust Legged Locomotion
[2022 ICRA] Learning Efficient and Robust Multi-Modal Quadruped Locomotion: A Hierarchical Approach
[2022 IROS] Robust High-Speed Running for Quadruped Robots via Deep Reinforcement Learning
[2023 RAL] Learning Robust and Agile Legged Locomotion Using Adversarial Motion Priors

Backflip
[2023 CoRL] Learning Agile Skills via Adversarial Imitation of Rough Partial Demonstrations
[2023 arXiv] Two-Stage Learning of Highly Dynamic Motions with Rigid and Articulated Soft Quadrupeds

Jump
[2021 CoRL] Learning to Jump from Pixels
[2021 ICRA] Learning Agile Locomotion Skills with a Mentor
[2022 TRO] Cat-like Jumping and Landing of Legged Robots in Low-gravity Using Deep Reinforcement Learning
[2023 RSS] Learning and Adapting Agile Locomotion Skills by Transferring Experience
[2023 RSS] Robust and Versatile Bipedal Jumping Control through Reinforcement Learning
[2023 arXiv] CAJun: Continuous Adaptive Jumping using a Learned Centroidal Controller
[2023 arXiv] Two-Stage Learning of Highly Dynamic Motions with Rigid and Articulated Soft Quadrupeds

Parkour
[2023 CoRL] Robot Parkour Learning
[2023 arXiv] Extreme Parkour with Legged Robots

Manipulation / Interaction
[2021 ICRA] Circus ANYmal: A Quadruped Learning Dexterous Manipulation with Its Limbs
[2022 ICRA] Hierarchical Reinforcement Learning for Precise Soccer Shooting Skills using a Quadrupedal Robot
[2022 arXiv] Creating a Dynamic Quadrupedal Robotic Goalkeeper with Reinforcement Learning
[2023 ICRA] DribbleBot: Dynamic Legged Manipulation in the Wild
[2023 ICRA] Legs as Manipulator: Pushing Quadrupedal Agility Beyond Locomotion

Pursuit-Evasion
[2020 IROS] Learning Agile Locomotion via Adversarial Training
[2023 arXiv] Learning Vision-based Pursuit-Evasion Robot Policies

Uncategorized
[2022 IROS] Advanced Skills by Learning Locomotion and Local Navigation End-to-End
[2023 ICRA] Advanced Skills through Multiple Adversarial Motion Priors in Reinforcement Learning

AMP: Adversarial Motion Prior

[2022 IROS] Adversarial Motion Priors Make Good Substitutes for Complex Reward Functions
[2023 CoRL] Learning Agile Skills via Adversarial Imitation of Rough Partial Demonstrations
[2023 ICRA] Advanced Skills through Multiple Adversarial Motion Priors in Reinforcement Learning
[2023 ICRA] Versatile Skill Control via Self-supervised Adversarial Imitation of Unlabeled Mixed Motions
[2023 RAL] Learning Robust and Agile Legged Locomotion Using Adversarial Motion Priors
[2023 arXiv] Learning Multiple Gaits within Latent Space for Quadruped Robots

State Estimation

[2022 RSS] Rapid Locomotion via Reinforcement Learning
[2022 RAL] Concurrent Training of a Control Policy and a State Estimator for Dynamic and Robust Legged Locomotion
[2023 ICRA] DreamWaQ: Learning Robust Quadrupedal Locomotion With Implicit Terrain Imagination via Deep Reinforcement Learning

Gait/Policy Transition

[2021 CoRL] Fast and Efficient Locomotion via Learned Gait Transitions
[2022 RAL] Learning Free Gait Transition for Quadruped Robots Via Phase-Guided Controller
[2023 ICRA] Expanding Versatility of Agile Locomotion through Policy Transitions Using Latent State Representation

Energy Efficiency

[2020 CoRL] Learning a Contact-Adaptive Controller for Robust, Efficient Legged Locomotion
[2021 TNNLS] Generic Neural Locomotion Control Framework for Legged Robots
[2021 CoRL] Fast and Efficient Locomotion via Learned Gait Transitions
[2021 CoRL] Minimizing Energy Consumption Leads to the Emergence of Gaits in Legged Robots
[2021 ICRA] Reinforcement Learning for Robust Parameterized Locomotion Control of Bipedal Robots
[2022 RAL] CPG-RL: Learning Central Pattern Generators for Quadruped Locomotion
[2022 RAL] Reinforcement Learning With Evolutionary Trajectory Generator: A General Approach for Quadrupedal Locomotion
[2022 TRO] RLOC: Terrain-Aware Legged Locomotion Using Reinforcement Learning and Optimal Control
[2022 ICRA] Learning Efficient and Robust Multi-Modal Quadruped Locomotion: A Hierarchical Approach
[2022 IROS] Robust High-Speed Running for Quadruped Robots via Deep Reinforcement Learning
[2022 IROS] Vision-Guided Quadrupedal Locomotion in the Wild with Multi-Modal Delay Randomization
[2023 CoRL] Deep Whole-Body Control: Learning a Unified Policy for Manipulation and Locomotion
[2023 CVPR] Neural Volumetric Memory for Visual Locomotion Control
[2023 RSS] Robust and Versatile Bipedal Jumping Control through Reinforcement Learning

Special Topics

Choice of action space
[2017 ACM SIGGRAPH] Learning Locomotion Skills Using DeepRL: Does the Choice of Action Space Matter?

Agility
[2018 TRO] Benchmarking Agility For Multi-legged Terrestrial Robots

Important sensory feedback
[2023 NMI] Identifying important sensory feedback for learning locomotion skills

Control frequency

[2023 ICRA] Learning Low-Frequency Motion Control for Robust and Dynamic Robot Locomotion

Emergence of gaits
[2021 CoRL] Minimizing Energy Consumption Leads to the Emergence of Gaits in Legged Robots

Fault-tolerant locomotion
[2021 IROS] Meta-Learning for Fast Adaptive Locomotion with Uncertainties in Environments and Robot Dynamics
[2023 RAL] FT-Net: Learning Failure Recovery and Fault-tolerant Locomotion for Quadruped Robots
[2021 arXiv] Reinforcement learning with adaptive curriculum dynamics randomization for fault-tolerant robot control*
[2022 arXiv] Saving the Limping: Fault-tolerant Quadruped Locomotion via Reinforcement Learning*

Low gravity
[2022 TRO] Cat-like Jumping and Landing of Legged Robots in Low-gravity Using Deep Reinforcement Learning

Torque control
[2023 arXiv] Learning Torque Control for Quadrupedal Locomotion

Risk-Aware/Risk-Averse
[2024 ICRA] Robust Quadrupedal Locomotion via Risk-Averse Policy Learning
[2024 ICRA] Learning Risk-Aware Quadrupedal Locomotion using Distributional Reinforcement Learning

[2017 ACM SIGGRAPH] Learning Locomotion Skills Using DeepRL: Does the Choice of Action Space Matter?
Peng 等人比较了四种不同动作（torques, muscle-activations, target joint angles, and target joint-angle velocities）在学习时间、策略鲁棒性、运动质量等方面的影响。[22]

[2018 RSS] Sim-to-Real: Learning Agile Locomotion For Quadruped Robots
Tan 等人提出了一套完整的供学习可控策略的训练系统。为了缩小仿真和现实的差距，通过对执行器进行精确的建模和延迟处理来提高物理仿真器的保真度，并通过两种步态（trotting，galloping）来衡量效果。【策略迁移】【系统辨识】[23]

[2018 TRO] Benchmarking Agility For Multi-legged Terrestrial Robots*
Eckert 等人提出了提出一种敏捷性（agility）基准测试，将敏捷性定义为以快速有效的方式执行一组不同特定任务的能力。在定义敏捷性之后，Eckert 等人定义了标准化基准值，度量方法以及用于敏捷性评分的在线数据库。[24]

[2019 TRO] Distributed Learning of Decentralized Control Policies for Articulated Mobile Robots*
Sartoretti 等人提出了利用asynchronous advantage actor-critic（A3C）算法的结构，将算法中的每个agent 定义为机器人每个独立可控的部分，例如将六足机器人的每条腿视为独立的agent。Sartoretti 等人展示了蛇形和六足机器人在非结构化地形中的闭环运动结果（分别为前进和保持机身平衡）。结果表明，该方法可以通过对机器人每个独立的部分进行分布式控制来适应许多不同类型的铰接式（articulated）机器人，并且可以以较高的样本效率训练分散策略。【分布式学习】[25]

[2019 CoRL] Data Efficient Reinforcement Learning for Legged Robots*
Yang 提出了一种基于模型的强化学习框架，在仅用4.5 min 时间收集数据就能实现四足机器人的行走。该框架结合了MPC 和RL 方法，在模型学习阶段，使用多步损失来降低误差累积，同时使用已习得的动力学模型所预测的未来状态执行动作以补偿规划延迟（planning latency），这使得学习的模型可以用于实时控制，并将足端轨迹的先验知识加入动作空间中以保证模型学习过程中的安全探索（动作的光滑性以及不损伤执行器）。该方法的样本利用效率比同期无模型方法高一个量级。为了提高强化学习训练中的数据有效性，yang等人使用基于模型的强化学习提升了数据有效性。在强化学习训练过程中交替执行模型训练和数据搜集过程。使用强化学习来调整MPC的预测从而提高精度，此外在学习期间引入了基于先验知识的步态规划器，提高了安全性。【MPC】【数据高效】[26]

[2019 ICRA] Realizing Learned Quadruped Locomotion Behaviors through Kinematic Motion Primitives*
Singla 等人认为人类和动物可以从很小的轨迹集中完成多种的任务，例如走路。基于这个想法，Singla 等人提出通过主成分分析从深度强化学习（D-RL）中学习的轨迹提取行走的基本运动模式，称为Kinematics Motion Primitives（kMPs），并且基于这些kMPs 直接重建关节轨迹，在四足机器人上实现{trot, walk, gallop, bound} 的行为。【KMP】[27]

[2019 ICRA] Trajectory-based Probabilistic Policy Gradient for Learning Locomotion Behaviors*
Choi 等人提出了基于轨迹的强化学习方法deep latent policy gradient （DLPG），将策略函数定义为关节轨迹上的条件概率分布（而不是给定状态输出动作），通过潜变量（latent variables）模型训练策略以实现较高的样本利用率。四足机器人Snapbot 通过该方法成功学习了前进和左右转向。【样本利用率】[28]

[2019 ICRA] Using Deep Reinforcement Learning to Learn High-Level Policies on the ATRIAS Biped
对于欠驱动的两足机器人，域随机化方法可能会使学习稳定的控制器变得更加困难。Li 等人提出学习一个神经网络策略，它作为结构化控制器的一部分，在仿真中结构化控制器的其他部分保持固定，且在必要时可以由专家进行调整。该方法在ATRIAS 上进行了验证，有助于加快仿真中的学习速度，且允许仿真到真实环境的迁移。[29]

[2021 RSS] RMA: Rapid Motor Adaptation for Legged Robots*
为了解决sim-to-real 的迁移问题，Kumar 等人提出了一种在线的系统辨识方法：Rapid Motor Adaptation （RMA）方法，环境因子编码（Env Factor Encoder）网络将一些环境配置变量编码

为一个潜向量，与运动策略一同进行训练；并构造适应模块（Adaptation Module）网络，采用监督学习的方式，以一系列历史状态信息估计作为输入预测上述编码网络的输出，从而实现对四足机器人对现实环境变化的快速适应。【策略迁移】【系统辨识】[1]

[2022 IROS] Adapting Rapid Motor Adaptation for Bipedal Robots*
双足机器人本质上比四足机器人更加不稳定，因此设计行走控制器是有难度的。Kumar 等人将原先RMA 的工作拓展到双足机器人的运动控制上并提出了Adapting RMA。在仿真训练的第一阶段中，一个用于编码环境因子的神经网络将与运动策略一同训练，训练过程使用了步态库生成一系列参考运动。在第二阶段，A-RMA 方法通过监督学习利用一系列历史信息估计环境编码网络的输出。与RMA 方法不同，A-RMA 引入了第三个阶段，即在仿真环境中使用上述不完美的估计器继续训练，让运动策略适应该估计器。实验中，Cassie 机器人的表现超过了基于RL 的控制器以及有模型的控制器。【策略迁移】[2]

[2023 CoRL] Deep Whole-Body Control: Learning a Unified Policy for Manipulation and Locomotion*
足式机械臂的标准模块化控制方法是将控制器解耦为运动和操纵两部分，但协调这两个模块是一个巨大的工程，并且在模块之间传递的误差将会导致不自然的动作。对此，Fu 等人提出了使用强化学习学习一个统一的全身控制器（WBC），并提出Regularized Online Adaptation （ROA）来缩小sim-to-real 的差距，以及利用动作空间中的因果依赖关系提出混合优势（Advantage Mixing）在训练初期减小信用分配（credit assignment）的复杂度，避免陷入局部极小值。ROA 是RMA 的变体：在同步训练策略网络和环境因子编码网络时，调整环境因子编码网络使之不会过分偏离Adaptation Module 的估计，并且通过在线模仿的方式训练Adaptation module。【机器人机械臂协作】【策略迁移】[3]

[2023 ICRA] Legs as Manipulator: Pushing Quadrupedal Agility Beyond Locomotion*
Cheng 等人希望能使足式机器人通过腿部完成一些基础的操作任务，例如学习爬墙，按按钮以及与物体进行交互。为此，他们将这些技能解耦为运动（行走、爬墙）和操纵（用一条腿与物体进行交互的同时用令三条腿保持平衡）两部分，并使用课程学习的方式完成训练。仅依赖机载传感器来学习爬墙这类技能是十分困难的，Cheng 等人提出regularized online adaptation （ROA）的变体来估计其他状态信息并达到优于Rapid Motor Adaptation （RMA）的表现。Cheng 等人还利用行为树（behavior tree）将这些技能结合到长期任务规划中。【组合策略】【课程学习】【策略迁移】[30]

[2020 IROS] Slope Handling for Quadruped Robots Using Deep Reinforcement Learning and Toe Trajectory Planning
Mastrogeorgiou 等人将深度强化学习与脚趾级别（toe-level）的轨迹规划相结合，由强化学习完成足端位置规划，轨迹规划模块输入足端位置输出脚趾位置，实现了在斜坡上小跑时，仍不会偏离目标，即机身偏航角是有界的。
【输出足端/脚趾轨迹】[31]

[2020 IROS] Learning Agile Locomotion via Adversarial Training
使用强化学习或者进化策略学习敏捷运动的主要挑战是设计可以诱导敏捷运动的环境和奖励函数。Tang 等人提出了可以促进敏捷运动形成的对抗学习算法，其中一个机器人（protagonist），学习追逐另一个机器人（adversary），而后者学习逃跑。当前最流行的多智能体算法（例如MADDPG，MATD3）并无法实现该目的。Tang 等人选择将问题解耦，并迭代训练追逐和逃跑行为，同时使用多组躲避策略防止protagonist 的过拟合。最终实验表明仅仅经过三代对抗训练，敏捷运动步态就会自动出现。【对抗学习】[32]

[2021 IROS] Rapidly Adaptable Legged Robots via Evolutionary Meta-Learning
为了设计适应高噪声环境的控制器，Song 等人引入抗噪声的Batch Hill-Climbing（BHC）适应算子，并将其与基于进化策略（evolutionary strategies）的元学习方法相结合，极大提高了机器人对高噪声的动力学变化的适应能力，使机器人能够根据不到3 分钟的真实数据调整其策略以适应变化。【抗噪声】【元学习】[33]

[2020 TNNLS] Teacher-Student Curriculum Learning
稀疏奖励任务是具有挑战的，其中一个解决该问题的方法是课程学习。Matiisen 等人提出了Teacher-Student Curriculum Learning （TSCL）框架– 学生尝试完成较复杂的任务时，教师算法挑选子任务给学生进行学习。基于学生应该多练习那些获得进步最快的任务的直觉，Matiisen 等人描述了一套教师算法。教师算法还通过选择学生表现越来越差的任务来解决遗忘问题。最终，在{addition of decimal numbers with LSTM, navigation in Minecraft} 两个任务上，该框架

取得与精心设计的课程相匹敌甚至更好的效果。【课程学习】[34]

[2020 RAL] MPC-Net: A First Principles Guided Policy Search
Version 1: Carius 等人提出了MPC-guided 的模仿学习方法MPC-net。该方法可以视为从完美critics（MPC）获取数据的策略迭代方法。其中核心的创新是基于最优控制第一原理的损失函数，即最小化control Hamiltonian。Carious 等人所提出的损失函数将最优化问题的约束条件进行了显式的编码，训练了混合专家（mixture-of-expert）神经网络用于控制机器人，并且成功不到10 分钟的示教数据中稳定各种不同步态。
Version 2: MPC-Net 是一种模仿学习方法，它使用MPC 的解决方案来指导策略搜索。主要思想是通过最小化控制哈密顿量来模拟MPC，同时通过参数化策略表示相应的控制输入。MPC-Net 可用于将模型预测控制器克隆到神经网络策略中。【MPC】[6]

[2020 RAL] Learning Fast Adaptation with Meta Strategy Optimization
足式机器人有能力在现实世界中的新场景下行走是一个很重要的议题。为此，Yu 等人提出了Meta Strategy Optimization （MSO），一种将潜在变量作为策略输入的元学习方法，其中核心思想是在训练和测试阶段接触相同的适应过程，即Strategy Optimization（SO），从而高效地学习运动技能以及适合快速适应的潜在空间（latent space）。【元学习】[35]

[2020 RAL] Guided Constrained Policy Optimization for Dynamic Quadrupedal Robot Locomotion
现有的强化学习方法不能保证机器人行为符合对真实场景下至关重要的必要安全约束，对此，Gangapurwala 等人提出了guided constrained policy optimization （GCPO）框架，该框架基于受限PPO 算法，让机器人遵循给定约束时跟踪速度指令。该框架比无约束RL 框架相比具有更快的收敛速度，且无需对奖励函数进行精细调整。【安全约束】[36]

[2021 CoRL] From Pixels to Legs: Hierarchical Learning of Quadruped Locomotion
当在控制回路中加入视觉输入时，足式机器人完成给定任务变得更加困难，因为它需要感知环境，同时处理腿部快速的运动。Jain 等人提出了一个分层框架，上层（HL）输入景深相机图片，输出潜指令（latent command），下层神经网络（LL Neural Network）接收HL 的潜指令，调节Trajectory Generator 的幅度和相位并输出TG，并且输出补充信号对轨迹进行调节。实验表明分层策略可以同时学习在悬崖和迷宫环境中进行运动和导航。【视觉融合】【分层网络】【步态规划器】【PMTG】[37]

[2020 CoRL] Learning a Contact-Adaptive Controller for Robust, Efficient Legged Locomotion
运动控制器可以呈现出在平地和崎岖路面都表现鲁棒的步态，但是很少有控制器能够为了适应环境变化改变步态，而具有环境适应性的步态会因减少不必要的运动而变得节能。Da 等人提出了一种结合基于模型的控制与强化学习的分层框架，该框架由一个上层控制器和一个底层控制器组成，前者指定足端的contact configuration，后者接收contact configuration 作为输入，通过QP 生成ground reaction forces（GRF），实现了比基准方法节能85%以及更鲁棒的效果。【MPC】【组合策略】【复杂地形】【分层框架】[17]

[2021 CoRL] Learning to Walk in the Real World with Minimal Human Effort
Ha 等人希望设计一个能使机器人在现实世界中自动学会行走的强化学习系统，其挑战主要来自两个方面：自动收集数据以及安全性。Ha 等人通过开发一个多任务学习步骤以及安全性约束的强化学习框架解决了该问题，实现只需要极少的人类干预就能让Minitaur 自动学习如何在各种地形{平地，软垫以及带裂纹的门垫} 上行走。【安全约束】【在线强化学习】[38]

[2021 ICRA] Sim-to-Real Learning of All Common Bipedal Gaits via Periodic Reward Composition*
足式机器人运动的一个关键挑战在于通过奖励函数描述并可靠地学习不同步态。Siekmann 等人提出了一个reward-specification 框架，将步态视为一系列周期性阶段，设计了基于基本的力和速度的概率周期成本，实现了{hop, gallop, run, walk} 多步态策略的学习。【策略迁移】[39]

[2021 ICRA] Reinforcement Learning for Robust Parameterized Locomotion Control of Bipedal Robots*
为两足机器人设计鲁棒的行走控制器是具有挑战性的。Li 等人结合了基于Hybrid Zero Dynamics（HZD）的步态库，提出无模型的强化学习框架，同时使用域随机化解决sim-to-real 的迁移问题，实现在真实的两足机器人Cassie 上的多种步行行为，例如跟踪目标速度，步行高度和转向偏航。【策略迁移】[40]

[2021 ICRA] Circus ANYmal: A Quadruped Learning Dexterous Manipulation with Its Limbs*
大多数四足的应用聚焦导航和巡检，但很少缺乏对环境的互动与操纵。Shi 等人采用无模型强化学习方法来训练深度策略，增加噪声将物理性质随机化，并在操作过程中添加随机干扰力，实现ANYmal 在没有任何接触式测量传感器的情况下使用其四肢稳健地平衡和操纵轻质球，所操纵的球最高旋转速度达到15° 每秒。【环境互动】[41]

[2021 IROS] Hierarchical Terrain-Aware Control for Quadrupedal Locomotion by Combining Deep Reinforcement Learning and Optimal Control
将对地形的感知与运动规划相结合是十分重要的。对此，Yao 等人提出了Hierarchical Terrain-aware Control（HTC）框架，将强化学习作为上层规划器，最优控制作为下层的控制器，其中地形高度图作为强化学习模块的视觉信息，决定了理想的落脚点和机身姿态，而最优控制负责计算站立腿的各关节力矩以维持身体平衡。最后通过台阶测试展示HTC 可以通过调节机身姿态有效地提高适应性。【分层网络】【MPC】【感知融合】【崎岖路面】[12]

[2021 IROS] Run Like a Dog: Learning Based Whole-Body Control Framework for Quadruped Gait Style Transfer*
为了使四足机器人能像真实世界中的狗一样运动，Yin 等人提出了基于学习的Whole-body Control（WBC）框架，基于多刚体动力学的底层控制器输出每个关节的期望力矩，而上层的神经网络规划期望步态和足端位置（foothold）。【分层网络】【WBC】[13]

[2021 IROS] Terrain-Aware Risk-Assessment-Network-Aided Deep Reinforcement Learning for Quadrupedal Locomotion in Tough Terrain
基于强化学习的控制策略在困难地形下运动时保证动作的稳定性仍存在难度。Zhang 等人提出集成了风险评估网络（RAN）的地形teacher-student 控制器来解决这个问题，其中RAN 通过评估历史或当前状态的风险等级作为惩罚项来指导策略更新，实现在仿真中可以穿越{Slippery Flat, Hills, Steps, Upward stairs, Downward stairs, Parapet} 等地形，以及在真实世界中trot 和bound。【teacher-student learning】【崎岖路面】[42]

[2021 IROS] Robust Feedback Motion Policy Design Using Reinforcement Learning on a 3D Digit Bipedal Robot*
针对两足机器人运动的策略鲁棒性问题，Castillo 等人提出一个分层框架，基于学习的上层策略网络输出参考轨迹，基于模型的底层反馈调控器输出参考轨迹的补充信号，实现只需要很小的调整就能实现sim-to-real 的迁移，并且在实际部署中对训练过程中并未出现的外力干扰和挑战性的地形仍能维持行走步态。【分层框架】[43]

[2021 IROS] Meta-Learning for Fast Adaptive Locomotion with Uncertainties in Environments and Robot Dynamics*
为实现机器人对不同变化的快速在线适应，从而产生多样且鲁棒的运动，Anne 等人提出基于模型的元强化学习运动控制策略，该方法不断更新交互（interaction）模型，对估计的状态-动作轨迹的可行动作序列进行采样，然后采用最优动作以实现奖励最大化。为了实现在线的模型自适应（model adaptation），Anne 等人所提出的Meta-adaptation 将根据过去0.2s 收集的样本选择潜向量，并学习每个训练条件对应的不同潜向量。该策略可以实现在极短时间内检测并适应环境变化，以及在光滑地面、外力扰动、电机故障、全腿截肢的情形下都展现鲁棒的运动。【元学习】[9]

[2021 arXiv] GLiDE: Generalizable Quadrupedal Locomotion in Diverse Environments with a Centroidal Model
Xie 等人提出了GLiDE 强化学习框架。基于此框架可以使用模型复杂度较低的中心动力学模型来训练强化学习模型并获得不亚于传统复杂物理模型的效果。[8]

[2021 CoRL] Minimizing Energy Consumption Leads to the Emergence of Gaits in Legged Robots*
Ver.1：预设步态限制了四足机器人在不同地形和不同速度下执行一般运动的能力。Fu 等人提出了基于蒸馏的学习方法以获得在不同速度下呈现自然步态转换的策略，并展示了在没有专家示教的情况下，能耗最小化对于产生自然的步态至关重要。
Ver.2：1981年，Hoyt 等人使用能耗最小化解释为什么动物在加速运动过程中转换步态。受此启发，Fu 等人采用“由合成进行分析”（analysis-by-synthesis）的方法，通过结合已训练好的的专家步态策略的强化学习奖励和蒸馏损失，在目标速度变化的情况下实现平稳的步态过渡，并且展示了能耗最小化对于不同速度下产生自然的步态具有重要影响。
Ver.3：Fu 等人在仿真中使用相同的超参数以及设计能耗奖励函数，为3 种目标速度训练了3 个单

独的策略- 0.375 m/s（低速）、0.9 m/s（中速）和1.5 m/s（高速）。在仿真中，以0.375 m/s 的目标速度下出现walk 步态，trot 步态在0.9 m/s 时出现，bouncing 步态在1.5 m/s 时出现。Fu 等人还使用了基于蒸馏的学习方法以获得在不同速度下呈现自然步态转换的（velocity-conditioned）策略。【步态转换】【RMA】[44]

[2022 CoRL] Visual-Locomotion: Learning to Walk on Complex Terrains with Vision*
训练机器人有效理解高维视觉输入是一个具有挑战性的问题。Yu 等人引入了一个分层框架，上层的基于视觉的上层策略将视觉信号和状态作为输入并输出期望立足点和机身运动，而这些运动通过底层的运动控制器完成，其中控制器包括控制摆动腿的位置控制器和控制站立腿的基于MPC 的力矩控制器。上层的视觉策略通过强化学习进行训练。该框架在包括{stepping stones, quincuncial piles, stairs, moving platforms} 的各种不平整路面上进行了验证。
Baseline：以PMTG 为架构的端到端神经网络策略，该策略直接输入深度图以及机器人状态（与Yu 等人提出的分层框架一致），输出理想关节角度。【分层网络】【MPC】【视觉融合】[45]

[2022 CoRL] Walk These Ways: Tuning Robot Control for Generalization with Multiplicity of Behavior*
Margolis 等人提出了一个策略学习框架，可以提高四足在out-of-distribution 情况下的表现。为了适应不同场景，Margolis 等人提出了Multiplicity of Behavior (MoB) 技巧，即在给定相同历史观测信息和一个行为参数集的情况下输出不同行为。当遇见新场景（out-of-distribution）时，可以通过改变行为参数来测试不同运动行为。[46]

[2022 CoRL] Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning
Rudin 等人提出了通过单块GPU 并行训练框架，并且提出了适合并行框架的课程学习方法。该并行方法可以在短短半小时内完成机器狗在平地或不平整地面上运动策略的训练。[47]

[2022 RAL] Neural Scene Representation for Locomotion on Structured Terrain*
在使用机载摄像机对机器人周遭环境进行深度测量时，摄像机的原生测量数据往往是有噪声的，且由于盲点的存在导致只能对周围环境进行部分观测。对此，Hoeller 等人提出一个三维重建模型，该模型由点云上的4D 全卷积网络和自回归反馈组成，其中神经网络从context 学习几何先验以完善场景。从深度传感器获得的带噪声点云和先前的输出一同输入网络，产生周围场景的点云估计，由于该方法的自回归特性，三维场景的重建将会被不断精炼，且网络可以记住那些进入盲点的区域（如果它们在过去已经被观测且重建）。【视觉感知】【表征学习】[48]

[2022 RAL] Concurrent Training of a Control Policy and a State Estimator for Dynamic and Robust Legged Locomotion*
Ji 等人提出了一个可以同时训练策略网络和状态估计网络的训练框架。前者输出期望关节位置，后者估计机器人的状态，例如机身速度，足端高度和足端触地概率。同时设计速度课程学习以实现高速运动。实验结果表明机器人可以穿越不同地形，平地上最高运动速度达到3.75m/s，在光滑平板（摩擦系数为0.22）上最高运动速度达到3.52m/s。Ji 等人提出了一种方法同时训练控制策略（强化学习）和状态估计器网络（监督学习，减小与真值误差）。训练完的整体框架可以实现实物机器人的快速运动和穿越不同地形。实验验证了机器人可以在斜坡和光滑地面上运动。亮点：对比试验设计问题：同时训练的好处？【同时训练】【状态估计】[49]

[2022 RAL] Energy-Based Legged Robots Terrain Traversability Modeling via Deep Inverse Reinforcement Learning
Gan 等人提出从原始本体传感器数据学习惯性特征，并将其结合到Terrain Traversability Model 中以提高模型保真度。为了解决足式机器人示教的次优性，Gan 等人提出在使用Maximum Entropy Deep Inverse Reinforcement Learning （MEDIRL）方法训练奖励网络的同时最小化轨迹排序损失（trajectory ranking loss），从而训练出更为节能的策略。【逆强化学习】[50]

[2022 RAL] Learning Free Gait Transition for Quadruped Robots Via Phase-Guided Controller*
Shao 等人提出了一个机器人在多种不同步态下进行运动的控制策略训练框架（Phase-Guided Reinforcement Learning），引入刻画四条腿的运动的四个独立相位，用于联系步态生成器和控制策略，从而将多步态学习任务转化为学习相位和足端运动之间的关系。【步态转换】[51]

[2022 RAL] Robust Predictive Control for Quadrupedal Locomotion: Learning to Close the Gap between Reduced- and Full-Order Models*
基于模板（template-based）的降阶（reduced-order）模型是一种实现实时轨迹规划的流行的方

法。为了弥补降阶模型和全阶模型的差距，Pandala 等人提出了一种基于凸二次规划（QP）的易于计算的鲁棒模型预测控制（RMPC）公式，并利用深度强化学习来训练神经网络来计算RMPC 框架的未被建模的动力学，实现了在不同地形上鲁棒且全盲的运动。
利用深度强化学习来训练一个神经网络来计算RMPC 框架的未建模动力学集合。最后通过大量的数值模拟和实验验证了所提出的控制器，用于A1 四足机器人在不同地形上的鲁棒和自主运动。【MPC】[19]

[2022 RAL] Learning to Navigate Sidewalks in Outdoor Environments*
Sorokin 等人希望让四足机器人遵循公共地图服务（Google 地图等）生成的路线行走，过程中保持在人行道上行走，并避免与障碍物和行人发生碰撞。受"learning by cheating"启发，他们设计了一个两阶段学习框架，首先在抽象世界中利用privileged information 训练teacher 网络，接着在仅使用真实传感器的情况下采用行为克隆将策略传授给student 网络。观测空间：Bird-Eye-View Image （BEV, privileged），Bird-Eye-View Lidar （BLID），Goal direction & distance（GDD）。动作空间：前进后退的速度，偏航转向。【导航】【teacher-student learning】[52]

[2023 ICRA] ViNL: Visual Navigation and Locomotion Over Obstacles *
目前在以前看不到的环境中进行室内视觉导航很大程度上局限于在地形平坦、整洁的家庭中使用轮式机器人。Kareer 等人希望实现足式机器人在（室内环境常见的）杂乱环境中的导航，为此，Kareer 提出Visual Navigation and Locomotion over obstacles （ViNL），包括基于视觉的导航策略（负责输出线速度和角速度指令的），以及基于视觉的运动策略（接收并跟随速度指令信号的同时控制关节防止机器人踩到障碍物）。两个策略分别在两个仿真器中进行训练，无需共同训练就可以一同部署到机器人上。仅利用机载视觉的ViNL 明显优于使用privileged 地形图的方法。【导航】[5]

[2023 CVPR] Neural Volumetric Memory for Visual Locomotion Control *
为了实现使用单个前置深度相机穿越复杂地形，并解决机器人必须依赖过去的观测推断当前机身下方的地形的问题，Kareer 等人提出Neural Volumetric Memory （NVM），一个对场景的3D 几何形状进行显式建模的几何记忆架构。NVM 聚合多个摄像机视角的feature volumes 并将它们融合成一个latent representation 用于运动控制，比朴素的帧堆叠方法效果更好。【复杂地形】[53]

[2022 IROS] Vision-Guided Quadrupedal Locomotion in the Wild with Multi-Modal Delay Randomization*
在具有各种障碍物、动态环境和不平坦地形的复杂环境中为四足机器人设计强大的视觉引导控制器具有很大难度，把仿真中训练好的视觉引导RL 策略部署到真实世界中是有挑战性的。Imai 等人认为由真实机器人不同组件中的不同延迟引起的异步多模态观测造成了sim-to-real 的巨大差距。为此，Imai 等人提出了Multi-Modal Delay Randomization （MMDR），通过在训练过程中将本体感知状态和视觉观测异步化来模拟现实世界中的异步性。实验证实机器人可以在避开障碍物的同时平稳地高速机动，比基准算法有显着改进。【策略迁移】[54]

[2022 IROS] Advanced Skills by Learning Locomotion and Local Navigation End-to-End*
Rudin 等人认为将导航任务拆分成子任务{path planning, path following, locomotion} 会限制机器人的能力，因为这些单独的任务并没有考虑到完整的解空间。Rudin 提出了一种基于强化学习的端到端策略训练方法，通过建立基于位置的任务奖励，即通过每个episode 结束时的最终位置进行奖励，同时增加惩罚项来实现sim-to-real 的迁移。该分层奖励结构要求所学策略先完成任务，再优化动作质量。该方法在ANYmal 上进行了验证，能够以更加节能的步态穿越先前无法穿越的各种崎岖地形。【策略迁移】【崎岖路面】【节能运动】[55]

[2022 IROS] Zero-Shot Retargeting of Learned Quadruped Locomotion Policies Using Hybrid Kinodynamic Model Predictive Control*
强化学习所学策略在另一个机器人上复用仍具有挑战，Li 等人结合Imitation RL 和MPC，提出了zero-shot policy retargeting 的框架，允许不同运动技能迁移到不同尺寸不同结构的机器人上，其中Imitation RL 用于生成轨迹和contact schedule，该轨迹被用来*seed* MPC，通过HKD 模型隐含地优化立足点位置。最终在不需要微调的前提下，成功将A1 和Laikago 的策略迁移到MIT Mini Cheetah 上。【模仿学习】【MPC】【策略迁移】[56]

[2022 IROS] Safe Reinforcement Learning for Legged Locomotion*
在现实世界中应用无模型的强化学习的一个主要瓶颈是安全性。基于该想法，Yang 等人提出了可以在完成目标任务的策略和防止机器人进入不安全状态的恢复策略之间进行转换的安全强化学习框架，当执行任务的策略违反安全约束时转换成恢复策略进行控制。该框架的有效性在仿真和现实中都得到了验证。其中在仿真环境下，机器人在{efficient gait, catwalk, two-leg balance, pace}

四个运动任务中比基准方法平均摔倒数减少了46.2%。【安全约束】[57]

[2022 IROS] Adversarial Motion Priors Make Good Substitutes for Complex Reward Functions*
人工设计的奖励函数往往需要繁琐的调参过程。Escontrela 等人以对抗性运动先验为基础，采用从参考运动数据集中学习样式奖励（style reward）的方法，证明训练策略的对抗性方法可以产生转移到真正的四足机器人的行为，而无设计需复杂的奖励函数。Escontrela 提出了一种可替代复杂奖励函数的"风格奖励"，它可以从运动捕捉演示的数据集中提取。风格奖励可以与任意的任务奖励函数相结合，训练使用自然策略执行任务的策略。这些策略使机器人能够从模拟中转移到现实世界中。他们对来自参考运动数据集的风格奖励进行编码–证明训练策略的对抗性方法可以产生转移到真正的四足机器人上的行为，而不需要复杂的奖励函数。他们还证明，有效的风格奖励可以从德国牧羊犬身上收集几秒钟的运动捕捉数据，从而产生一个自然的步态过渡，以达到节能的运动策略。【对抗学习】【简单奖励】[58]

[2022 ICRA] Hierarchical Reinforcement Learning for Precise Soccer Shooting Skills using a Quadrupedal Robot*
构造让四足机器人向给定目标踢足球的任务颇具挑战性，同时包含了机器人的运动控制以及规划。该任务具备多个难点：例如动力学限制、运动稳定性以及如何将难以建模的可形变足球踢向某个给定目标的运动规划。Ji 等人采用分层框架来解决这种问题。首先训练可以在站立状态下既能用一个足端跟踪任意轨迹，同时保持平衡的控制策略。接着，训练机器狗将足球提到期望位置。在仿真中，Ji 使用刚性球进行训练，在真实世界中进行部署时需要进行微调。【环境互动】【分层网络】[59]

[2022 arXiv] Creating a Dynamic Quadrupedal Robotic Goalkeeper with Reinforcement Learning*
足球守门是一个富有挑战性的问题，结合了高动态的运动（拦截动作往往在不到一秒钟就完成了）和精确而迅速的物体操纵。Huang 等人通过一个分层的无模型强化学习框架解决了该问题，其中第一部分包含多个底层控制策略，每个策略可以实现不同的运动技能，同时足端还可以跟踪参数化轨迹，第二部分是用于选择理想运动技能和上层规划器，最终实现了在真实环境中87.5%的拦截成功率。【环境互动】【分层网络】[60]

[2022 IET]
Li 等人希望能够在四足机器人上实现跳跃步态并适应不同地形，为此Li 等人先利用基于模型的控制器收集跳跃数据并预训练网络，接着通过强化学习对上述网络的权重进行优化。Li 设计了一个考虑触地状态的奖励函数用于鼓励步态的对称性和周期性。[61]

[2022 ICRA] Legged Robots that Keep on Learning: Fine-Tuning Locomotion Policies in the Real World*
足式机器人可以通过在多种环境中进行强化学习，从而获得鲁棒的控制策略，但是这种训练方式很难预测现实世界中可能遭遇的所有情形。基于让机器人在所处任何环境中持续学习的想法，Smith 等人提出了一个在现实环境中对运动策略进行微调的强化学习框架，实现仅需现实世界中适当的训练，就可以极大地提高运动表现。【真实环境训练】[62]

[2022 TRO] RLOC: Terrain-Aware Legged Locomotion Using Reinforcement Learning and Optimal Control*
Gangapurwala 等人提出了一种统一的基于模型和数据驱动的四足规划和控制方法，以实现在不平整地形上的运动。利用机载本体感知和外部感知反馈，使用强化学习策略将传感信息和速度指令映射到足端规划中。Gangapurwala 还引入了两个辅助强化学习策略用于矫正全身运动和恢复控制（recovery control），以应对物理参数的变化和外部扰动。最终在各种复杂地形上验证了算法的鲁棒性，并展示了可迁移性。【MPC】【感知融合】[18]

[2023 CoRL] DayDreamer: World Models for Physical Robot Learning*
Dreamer 算法通过习得的world model 中进行规划，从少量交互中学习，最近在视频游戏中的表现优于纯强化学习。学习world model 以预测潜在动作的后果能够在假想中进行规划，减少试错的数量。Wu 等人提出将Dreamer 算法应用于真实的机器人，从而实现无需仿真器的在线强化学习。【真实环境训练】[63]

[2023 CoRL] GenLoco: Generalized Locomotion Controllers for Quadrupedal Robots*
当前大多数基于学习的控制器开发框架都专注于训练特定机器人的控制器，对于每个新机器人都需要重复该训练过程。为此，Feng 等人提出一种可以程序化的生成一组用于训练的仿真机器狗的形态随机化方法，合成的通用运动控制器可直接部署到具有相似形态的各种四足机器人上。【策

略迁移】[64]

[2023 CoRL] Learning Agile Skills via Adversarial Imitation of Rough Partial Demonstrations*
通过强化学习学习敏捷的技能往往需要需要在奖励励函数中提供显式的任务信息。Li 等人提出了Wasserstein Adversarial Behavior Imitation (WASABI)，从粗糙且仅包含部分信息的演示中推断奖励函数，以便在不容易获得参考或专家演示的情况下成功获得技能。Li 等人在Solo 8 上测试了后空翻技能，该技能忠实地复制了手持式的人类演示（机身在手持式后空翻的示教）。【生成对抗】[65]

[2023 ICRA] OPT-Mimic: Imitation of Optimized Trajectories for Dynamic Quadruped Behaviors*
许多研究作使用动捕数据或人工生成的轨迹作为参考运动，但很少有工作探索使用基于模型的轨迹优化产生的参考运动。为此，Fuchioka 等人提出基于轨迹优化和模仿许的运动生成框架，轨迹优化用于生成适用于简化单刚体（SRB）模型的开环轨迹，然后通过基于模仿的强化学习（RL）对其进行跟踪，以生成能够在全阶模型上执行这些运动的闭环反馈控制器。然后，这些运动可以直接部署在物理机器人上，而无需额外的在线模型适应（model adaptation）。【轨迹优化】【模仿学习】[66]

[2023 ICRA] Advanced Skills through Multiple Adversarial Motion Priors in Reinforcement Learning*
强化学习的一大挑战是为实现所需运动风格的繁琐调参过程。对此，Vollenweider 等人提出了一种基于对抗运动先验的强化学习方法Multi-AMP，允许多种离散可切换的运动风格，以实现更高阶的技能（在四足与人形之间进行切换）。Vollenweider 等人在轮式四足机器人上进行了实验，其中包括从现有的RL 控制器和轨迹优化中学习技能，例如躲避和行走，以及在四足和人形之间切换等新技能。【AMP】[67]

[2023 ICRA] Versatile Skill Control via Self-supervised Adversarial Imitation of Unlabeled Mixed Motions*
给定包含不同行为的无标注运动片段的大型数据集，提取和学习技能是具有挑战性的。为此，Li 等人提出了Cooperative Adversarial Self-supervised Skill Imitation (CASSI)，在生成对抗模仿学习框架中应用无监督技能判别，从包含各种无标注运动的参考数据集中提取并学习高级技能。该策略在Solo 8 机器人上进行了部署和测试而无需进一步调整（adaptation），并忠实地模仿了示教的多种技能。【模仿学习】【生成对抗模仿学习】[68]

[2023 ICRA] Expanding Versatility of Agile Locomotion through Policy Transitions Using Latent State Representation*
融合多个执行特定步态的策略需要一个鲁棒的策略转换方法。Christmann 等人通过策略的潜在状态表示（latent state representation）编码机器人的状态和一些环境因素，采用基于学习的方法设计了用于辨识转换能否成功的网络transition-net，允许在不改变当前策略的情况下拓展策略库。该方法在真实环境中的{stand, pace, trot} 的策略库下进行对比验证。【组合策略】[69]

[2023 ICRA] Learning to Walk by Steering: Perceptive Quadrupedal Locomotion in Dynamic Environments*
为了解决动态环境中的视觉运动问题，Seo 等人提出了一个分层学习框架PRELUDE，将视觉运动问题分解为上层的导航指令预测以及底层的步态规划器。前者通过人类示教的模仿学习实现，后者通过强化学习训练得到。该方法可以实现复杂的导航行为。【感知融合】【模仿学习】【分层网络】[70]

[2023 ICRA] DreamWaQ: Learning Robust Quadrupedal Locomotion With Implicit Terrain Imagination via Deep Reinforcement Learning*
目前最先进的RL 方法依赖复杂可靠的感知框架。此前只依赖自身传感器的工作很少展示具备长距离穿越具有挑战性的地形的能力。Nahrendra 等人提出一个的学习框架，可以仅靠自身传感器（proprioception）隐式推断地形属性，使得四足机器人即使在有限传感器条件下仍然可以习得穿越具有挑战性地形（如台阶）的策略。Nahrendra 等人构造了环境辅助估计器网络（context-aided estimator network）同时估计机身状态以及环境。该框架允许四足机器人穿过具有挑战性的地形。即使是在有限的传感方式下，也能在具有挑战性的地形中行走。所提出的框架在现实世界的户外环境中得到了验证。在长距离路程中的不同条件下都验证了该框架的可行性。【状态估计器】【崎岖路面】[71]

[2023 ICRA] Force control for Robust Quadruped Locomotion: A Linear Policy Ap-

proach*
Shirwatkar 等人提出基于质心动力学的MPC方法用于生成参考轨迹数据，然后使用模仿学习训练线性策略，以最小化与参考轨迹的偏差。该控制器计算效率高，并且在仿真和现实中，在室内和室外地形中对推力都具备恢复能力。【模仿学习】[10]

[2023 ICRA] Optimizing Bipedal Locomotion for The 100m Dash With Comparison to Human Running*
Crowley 等人结合LSTM 网络以及动力学随机化（dynamics randomization）的训练框架。在该训练框架下，可以通过适当规范奖励函数在各种速度范围内优化不同的步态。Crowley 等人还将优化后的跑步步态集成到一个完整的控制器中，该控制器满足100 米短跑的规则。部署后的Cassie 打破了两足机器人100m 短跑的吉尼斯世界纪录。【域随机化】[72]

[2023 ICRA] Learning Low-Frequency Motion Control for Robust and Dynamic Robot Locomotion*
传统观念认为可以增加运动控制频率来实现运动鲁棒性和反应性最大化，然而Gangapurwala 等人通过学习一个控制频率为8 Hz 的策略并在ANYmal C 上实现鲁棒运动挑战了该观念。同时他们用控制频率从5 Hz 到200 Hz 的运动策略做了对比分析，并发现低频率控制策略对执行器作用延迟和系统参数变化较为不敏感。[73]

[2023 ICRA] Efficient Learning of Locomotion Skills through the Discovery of Diverse Environmental Trajectory Generator Priors
传统PMTG 架构采用单一TG，但是单一TG 是否能够很好地应对多种地形仍存在疑虑。一种可能性是策略需要通过大量工作弥补单一TG 的不足。Surana 等人提出了MAP-Elites（一种Quality-Diversity 算法）和PMTG 架构的强化学习方法。其中MAP-Elites 算法负责找到以TG 形式的多样化先验，每个先验都是针对相应的任务和环境而专门设计的且表现较好的。TG 的参数空间将作为QD 算法搜索的解空间。在每次训练的episode 中，首先会选择目标环境和TG 参数，接着强化学习策略调节TG 信号并输出补充信号。[74]

[2023 RSS] Learning and Adapting Agile Locomotion Skills by Transferring Experience*
为机器人设计鲁棒的控制器以完成高度敏捷的运动仍然是一个重大挑战。为此，Smith 等人提出了从现有控制器转移经验，以快速开始学习新任务的框架。现有的控制器可能已针对不同动态下的不同目标进行了优化，或者可能需要对周围环境的不同了解，因此对于目标任务而言可能非常欠佳，而Smith 等人所设计框架允许从这次欠佳控制器中进行灵活学习，且实现了复杂的敏捷跳跃行为以及用后腿行走时导航到目标位置的运动技能，并能适应新环境。【模仿学习】【策略迁移】[75]

[2023 RSS] Robust and Versatile Bipedal Jumping Control through Reinforcement Learning*
Li 等人希望能让双足机器人完成各种各样的跳跃任务，例如跳到不同的位置和方向。为了，他们提出了新的策略结构，将长期Input/Output（IO）历史进行编码，与此同时直接提供短期的IO 历史，以提高在这些任务上的表现，并使用了多阶段训练方法，在Cassie 上实现真实世界中各种具有挑战性的跳跃任务，例如立定跳远、跳上高架平台和多轴跳跃。【分段训练】[76]

[2023 RAL] Learning Robust and Agile Legged Locomotion Using Adversarial Motion Priors*
Wu 等人提出第一个盲运动系统，将基于Trajectory Optimization（TO）生成的对抗运动先验Adversarial Motion Priors（AMP）数据集与teacher-student 训练框架相结合，在仅使用本体传感器的情况下，实现在自然地形上快速移动时稳健地穿越复杂地形。teacher policy：可以接触到本体传感器信息，privileged 状态信息以及地形信息。其奖励函数包括任务奖励，风格奖励和（用于约束运动的）规则化奖励。student policy；通过监督学习的方式模仿teacher 网络输出的动作。在仅使用本身传感器的情况下重构teacher 策略中（由privileged 状态信息以及地形信息编码）的潜表征向量。【复杂地形】【teacher-student learning】【AMP】[77]

[2023 RAL] Learning-Based Design and Control for Quadrupedal Robots With Parallel-Elastic Actuators
并行弹性关节可以通过辅助执行器增加扭矩来提高机器人的效率和强度。Bjelonic 等人提出一个设计优化框架，使得并行弹性膝关节可以和运动控制器被一同优化。在第一阶段，通过强化学习训练能适应一个既定的设计参数集的运动策略，在第二阶段，通过贝叶斯优化选择最优设计。该设计能使机器狗在不影响跟踪性能的情况下更加高效输出力矩。[78]

[2023 RAL] Learning Complex Motor Skills for Legged Robot Fall Recovery
Yang 等人提出通过选择关键状态进行初始化（KSI）学习摔倒之后的站立恢复策略，并将该方法与Random State Initialization（RSI）进行比较。该方法的有效性在各种不同尺寸、构型的双足和四足机器人上进行了验证，实验证实KSI 在训练四足机器人恢复策略时收敛速度更快，同时在训练双足机器人时最终奖励更高。[79]

[2023 RAL] FT-Net: Learning Failure Recovery and Fault-tolerant Locomotion for Quadruped Robots
在足式机器人的工作过程中，随时可能发生各种严重的硬件故障。Luo 等人提出了故障恢复和故障容错控制框架FT-Net，该框架采用RMA 架构，将参数化的不同故障场景进行隐式编码，并通过adaptor 统一调节故障情况下的运动控制策略。同时Luo 等人还采用了启发式的奖励函数项，其中包含与机器人动力学（VHIP）相关的设计。[4]

（2023 NMI）Identifying important sensory feedback for learning locomotion skills
在机器人学习中，了解不同传感器反馈定量的相对重要性对于产生所需行为的学习方法至关重要，而该领域仍缺乏相关研究。Yu 等人提出了可以定量评估强化学习中不同传感器状态反馈的相对重要性的显著性分析（saliency analysis），并对一个特定任务下各个传感器反馈的相对重要性进行排序。Yu 证实仅使用最基本的传感器反馈，包括关节位置、重力向量以及机身线速度和角速度所训练的策略，其表现足以匹敌那些利用更多反馈所训练的策略。[80]

[2023 arXiv] Learning Quadruped Locomotion using Bio-Inspired Neural Networks with Intrinsic Rhythmicity
在结合CPG 和MLP 网络的训练框架下，由于CPG 网络的内在周期性，在采用基于梯度的学习进行训练时，需要时间反向传播（BPTT）来获得CPG 的参数。Yang 等人提出将CPG 的隐藏状态显式作为网络的输入和输出，从而转变为完全可微分的无状态网络，采用DRL 同时优化MLP 和CPG 的参数。训练好的网络可以在仿真环境中通过不平整路面和抵抗外力。Yang 等人还讨论了网络的可解释性。【CPG】[81]

[2023 arXiv] RL + Model-based Control: Using On-demand Optimal Control to Learn Versatile Legged Locomotion*
Kang 等人提出了一种结合model-based 的最优控制和无模型的强化学习的控制框架以完成多种鲁棒的运动，通过RL 模仿使用最佳控制框架生成的（覆盖一定范围速度和步态的）参考运动来实现motion imitation，以及通过考虑全身动力学（Whole-body Dynamic）克服简化模型的固有局限性。习得的策略成功地产生了多种的步态，能有效地外力以及穿越崎岖和倾斜的地形。【最优控制】[11]

[2023 arXiv] Two-Stage Learning of Highly Dynamic Motions with Rigid and Articulated Soft Quadrupeds*
许多工作认为引入弹性可以产生更好的表现，然而，这也为控制带来了额外的挑战。一些基于RL 的工作将被动弹性（passive elasticity）应用到四足运动，但不是应用到该工作所展示的杂技动作上。Vezzi 等人提出结合强化学习和进化算法的两阶段策略以学习高动态的运动技能。在进化阶段，无需预先接触专家演示的情况下直接学习满足任务需求的关节指令。该阶段使用augmented random search (ARS) 算法学习一个线性策略。在第二阶段，通过模仿学习的方法对更复杂的策略神经网络进行热启动，接着使用强化学习的方式改进策略网络。该训练策略在仿真环境中进行了验证以及对比实验，其中包括完成{jumping, pronking, back-flip} 等杂技任务。【分段训练】【进化算法】【模仿学习】[82]

[2023 arXiv] Learning Multiple Gaits within Latent Space for Quadruped Robots
动物已经进化出在不同地形和速度下一系列鲁棒而高效的步态，但是在多足机器人上复制这些自然的步态转换是十分困难的。Wu 等人提出由步态编码器和步态生成器共同构成的潜空间（latent space）以重用多种步态技能，实现可适应性（adaptive）的步态行为。同时Wu 等人还设计了一个步态依赖的奖励函数，该奖励函数由显式的步态参数和隐式的Conditional Adversarial Motion Prior（CAMP）构成。该控制器在Go1 上进行了部署，成功复制了与CAMP 数据集相似的的自然步态。[83]

[2023 CoRL] Legged Locomotion in Challenging Terrains using Egocentric Vision*
为了利用机载视觉和传感器信息实现复杂地形上的运动，Agarwal 等人提出两阶段训练方法，在第一阶段采用低精度高程图（elevation map）作为privileged 信息，第二阶段将第一阶段的策略进行蒸馏，不同之处是输入由高程图变成了深度图。Agarwal 等人还采用了两种架构，一是统一的RNN 架构（在第二阶段中，使用第一阶段的策略网络作为初始化网络并继续训练该网络），二是解耦的RMA（Rapid Motor Adaptation）架构（在第二阶段固定策略网络）。[84]

[2023 CoRL] Robot Parkour Learning*
为实现自主的复杂跑酷系统，包括匍匐前进、翻越障碍物、跳跃间隔、穿越狭缝等，Zhuang 等人提出了基于视觉的两阶段RL 训练框架：在预训练阶段允许机器人“穿透”障碍物，采用实行soft dynamic constraints 的自动课程学习，在精调阶段实行所有dynamic constraints。在学习了每个跑酷技能后，Zhuang 等人使用DAgger 将这些技能蒸馏成一个基于视觉的跑酷策略。【Parkour】[85]

[2024 ICRA] Extreme Parkour with Legged Robots*
Cheng 等人希望在四足机器人上实现跑酷技能（跳上箱子，跳跃间隔、斜坡等），为此他们采用两阶段teacher-student 训练方法。第一阶段，先利用RL 学习一种运动策略，该过程可以访问一些privileged information（环境参数和扫描点（scandots）以及作为大致方向引导的标志点），接着采用Regularized Online Adaptation（ROA），即利用历史观测对环境信息进行推断。第二阶段，从scandots 提取策略，该策略网络直接以机载视觉图像和机载传感器作为输入，并输出关节位置信号。此外机器人还能根据障碍物自行调整行进方向（yaw）。【Parkour】[86]

[2024 ICRA] Learning to walk in confined spaces using 3D representation*
机器人穿越狭小空间仍是一个挑战，特别是传统2.5D Lidar 数据无法捕捉悬空物体的特征，Miki 等人提出使用强化学习和3D volumetric representations，并采用分层策略结构，使底层鲁棒策略能够跟踪6 维指令，上层策略能够具有空间感知，并在有悬空物体的情况下进行导航。【Lidar】[87]

[2024 ICRA] Robust Quadrupedal Locomotion via Risk-Averse Policy Learning*
机器人运动过程中总是会遇到很多不确定性，例如突然的环境变化或者外力冲击。使用分量回归(quantile regression) 学到的分布价值函数(distributional value function) 来模拟环境的偶然不确定性，并通过风险扭曲度量(risk distortion measure) 优化最坏情况下的场景来执行倾向于规避风险的策略。【风险意识】【盲运动】[88]

[2024 ICRA] Learning Risk-Aware Quadrupedal Locomotion using Distributional Reinforcement Learning*
与Risk-Averse Policy Learning 所采用的盲运动不同，Schneider 等人采用了Distributional Proximal Policy Optimization (DPPO)，让具备外部传感器的四足机器人学习的可以选择不同风险偏好程度的策略。【风险意识】【外部感知】[89]

[2023 RAL] Imitating and Finetuning Model Predictive Control for Robust and Symmetric Quadrupedal Locomotion*
为了同时发挥MPC 和强化学习的优势，Youm 等人提出了Imitating and Finetuning Model Predictive Control (IFM) 框架。该框架分为三个阶段: stage-0 输出MPC 力矩，并将该力矩转化为期望关节角度；stage-1 为模仿学习，策略网络需要模仿MPC 输出的动作；stage-2 采用强化学习方法对所学策略网络进行微调。【模仿学习】【MPC】[90]

[2024 RAL] MorAL: Learning Morphologically Adaptive Locomotion Controller for Quadrupedal Robots on Challenging Terrains*

【形态适应】[91]

[2024] Deep Compliant Control for Legged Robots*
Hartman 等人引入了一个显式的恢复阶段，其中跟踪奖励是给定的，与控制策略产生的运动无关。【柔顺控制】[92]

[2022 RAL] CPG-RL: Learning Central Pattern Generators for Quadruped Locomotion*
针对CPG的参数难以整定及动态调节的问题，Bellegarda提出了一种CPG与RL融合的方法：CPG-RL，它使用RL动态调节CPG的参数，实现了可变机身高度和足端高度的四足运动，此外，这种方式仅需要观测触地状态就可以实现机身的前向运动。方法展现出了很强的鲁棒性。【CPG】【Ijspeert】[93]

[2022 Nature SR] Controlling the Solo12 quadruped robot with deep reinforcement learning*
Aractingi等人在Solo12上通过深度强化学习训练了一个端到端控制器，控制器输出各关节的参考运动，实现了机器人在不同地面上的运动，包括崎岖路面。【崎岖路面】[94]

[2022 NMI] High-speed quadrupedal locomotion by imitation-relaxation reinforcement learning*
Jin等人提出了一种模仿-松弛的强化学习方案（imitation relaxation reinforcement learning，IRRL）实现了机器人的高速运动，这种方案可以分阶段去进行强化学习，第一阶段是基于模仿的奖励函数设计，第二阶段是基于效果的奖励函数设计。文章引入了一种基于随机稳定性的方法进行系统鲁棒性的分析，并且发现状态空间熵的降低是一个量化指标，可以检测到潜在的混沌和分叉。【模仿学习】【分段学习】[95]

[2022 RAL] Reinforcement Learning With Evolutionary Trajectory Generator: A General Approach for Quadrupedal Locomotion*
Shi等人提出了进化的基于CPG的轨迹生成器与RL相结合的方法，交替使用黑箱算法训练轨迹生成器，使用强化学习训练补充信号，在简单的奖励函数下实现了机器人的多模态任务控制，包括上下楼梯，过平衡木等。【步态规划器】【交替训练】[96]

[2022 RAL] DeepGait: Planning and Control of Quadrupedal Gaits Using Deep Reinforcement Learning*
Tsounis等人提出了一种基于强化学习的方案进行崎岖路面的运动DeepGait。这种方法结合了基于模型的运动规划和强化学习。结果显示机器人可以穿越台阶，缺口和独木桥。【MPC】【崎岖路面】[97]

[2019 IROS] Hierarchical Reinforcement Learning for Quadruped Locomotion
Jain等使用一种分层式的强化学习方式来自动分解复杂的控制任务。一个频率低的高层的策略接收上层指令，并且选择下层策略及其执行时间。这种方法可以实现机器人的路径跟随及转向。【分层网络】【组合策略】[98]

[2021 ICRA] Real-Time Trajectory Adaptation for Quadrupedal Locomotion using Deep Reinforcement Learning*
Gangapurwala等人提出了一种架构，使用强化学习去生成一个相对于参考轨迹的偏移，这个参考轨迹是由检测地形的轨迹优化器生成的，强化学习接收经过编码后的高度图信息和生成的轨迹，及当前位置。最终的修正位置输入到whole body controller之中。【感知融合】【崎岖路面】【WBC】[15]

[2022 IROS] Learning Coordinated Terrain-Adaptive Locomotion by Imitating a Centroidal Dynamics Planner*
为了解决在崎岖地面上寻找精确落脚点的问题，Brakel等人使用了一种组合了轨迹优化和学习的方式，实现了可以适应地形的控制器。训练策略模仿非线性求解器求解出的轨迹可以将该策略应用到没有见过的地形上，标准强化学习难以做到这一点。【MPC】【崎岖路面】[99]

[2022 IROS] Robust High-Speed Running for Quadruped Robots via Deep Reinforcement Learning*
为了实现机器人稳定的高速运动，Bellegarda等人提出了一个学习框架，直接在任务空间（笛卡尔空间）中选择动作，这种方式相较于在关节空间中选择动作，提高了训练的速度和样本效率，并且使用简单的奖励函数就可以实现自然的步态。实验验证了sim2sim和sim2real的效果。【输出足端轨迹】[100]

[2020 RSS] Learning Agile Robotic Locomotion Skills by Imitating Animals*
为了再现动物的敏捷移动技能，Peng等人提出了一个模仿动物运动的框架，使用强化学习的方式

利用参考运动数据，使得四足机器人模仿参考数据实现了敏捷运动，包括跳跃，不同步态行走等。学习的策略可以自适应环境，通过领域适配的技术来减小策略迁移时的性能损失。【模仿学习】【策略迁移】[101]

[2021 IROS] A Hierarchical Framework for Quadruped Locomotion Based on Reinforcement Learning*
为了解决四足机器人控制器难以训练和难以部署的问题，Tan等人提出了一个分层式的强化学习网络，上层策略接收机身输出下层步态规划器的参数，下层步态规划器生成动态的足端轨迹。【崎岖路面】【步态规划器】[102]

[2020 ACMSG] ALLSTEPS: Curriculum-driven Learning of Stepping Stone Skills*
为了解决人形机器运动时选择落脚点的问题，Xie等人提出了一种基于课程学习的强化学习架构，实验证明了课程学习对于高效进行强化学习的重要性。【课程学习】[103]

[2022 TRO] Cat-like Jumping and Landing of Legged Robots in Low-gravity Using Deep Reinforcement Learning*
为了解决在低重力环境下的跳跃和着陆的问题，rudin等人使用强化学习训练了一个神经网络控制器，完成了足式机器人跳跃时的姿态控制。实现了在三维的姿态重定位和着陆姿态的控制。RL可以端到端实现实现跳跃中的2D和3D重定位，和落地时的接触动力学设计。MPC可以部分实现其中的子任务，但是难以整体实现全部的跳跃和落地过程。【跳跃运动】[104]

[2021 RAL] CPG-Based Hierarchical Locomotion Control for Modular Quadrupedal Robots Using Deep Reinforcement Learning
为了解决模块化四足机器人的自适应运动控制问题，wang等人使用底层CPG和上层RL构建控制器，CPG进行预训练生成稳定的步态，上层RL根据状态和指令调节CPG参数。通过很少的已知知识和高效学习机器人可以实现给定运动，包括不平的地面和外界扰动。【CPG】【崎岖路面】[105]

[2021 ICRA] Learning Agile Locomotion Skills with a Mentor*
为了解决强化学习方法实现四足机器人在复杂地形中敏捷运动，（跳过缝隙和跨栏）其奖励难以设计及复杂的课程设计问题，Iscen等人使用了一种多阶段（3阶段）方式去训练控制器，第一阶段，使用一个简单的任务去训练学生及找到任务对应最好的mentor，第二阶段则注重于普适性，通过不同和逐渐变难的环境去增加控制器的效果，第三阶段则逐渐减少mentor的介入。机器人在该方法下在仿真中可以穿越裂缝和跨栏。【跳跃运动】【分段学习】[106]

[2022 ICRA] Learning Efficient and Robust Multi-Modal Quadruped Locomotion: A Hierarchical Approach
为了解决强化学习方法实现四足机器人多模态运动的问题，xu等人提出了一种分层式的设计方法，上层网络用于步态选择及输出自适应模型参数，底层算法用于实现鲁棒的步态控制（MPC）。【分层控制】【MPC】[21]

[2019 SCI RO] Learning agile and dynamic motor skills for legged robots*
为了实现基于学习的足式机器人精准和有效的电机控制，hwangbo等人提出了使用真实数据训练电机执行器神经网络模型提高RL仿真数据精度的方法，在该方法下训练出来的策略可以直接部署到实物机器人上，且可以实现精准且节能的运动。【系统辨识】【策略迁移】[107]

[2020 SCI RO] Multi-expert learning of adaptive legged locomotion*
为了实现多功能的四足机器人运动，yang等人提出了一种多专家学习架构MELA（Multi-expert learning of architecture）。首先预训练多个DNN网络（专家）用于实现不同的任务，再训练一个门控神经网络（GNN）学习这些DNN 的组合。MELA可以实现不同运动模式及其过渡，并动态生成新的DNN网络。实验验证了四足机器人可以自主实现连贯的运动，转向和跌倒回复。【组合策略】[108]

[2023 SCI RO] Learning quadrupedal locomotion on deformable terrain*
为了解决在基于强化学习方法下在可变形地面上的高速运动问题，Choi等人通过建立了一种颗粒模型用于构建强化学习的环境，并引入了一种自适应的控制方案去辨识地形的属性。这个工作在Raibo机器人上进行了实现，实验证明可以进行高速运动。【基于模型】【崎岖地面】[109]

[2021 ICRA] Imitation learning from mpc for quadrupedal multi-gait control*
为了实现在单一策略下生成多种步态，reske等人使用模仿学习方法，扩展了MPC-net，使用混合专家网络mixture-of-experts network（MEN），文章中提出了一种新的奖励函数来选择不同的专

家策略。实验证明，相对于MPC和行为克隆，文章提出的方案在崎岖路面下具有优势。【模仿学习】【MPC】[7]

[2018 CoRL] Policies modulating trajectory generators
Iscen等人提出了一种策略调制轨迹生成器（Policies Modulate Trajectory Generators，PMTG），它可以参数化生成周期性运动。使用深度强化学习可以对其进行训练，仅需IMU观测就可以在快速训练下实现速度控制。文章开展了实物实验,在自制四足机器人上进行了验证。【轨迹生成器】[110]

[2020 ICRA] Learning generalizable locomotion skills with hierarchical reinforcement learning*
为了提高训练的效率和策略的泛化性，Li等人提出了一种分层式强化学习，将策略分层了两层，上层使用基于模型的规划选择不同的运动primitive，而下层则包含了一组不同的训练得到的primitives，可以生成各关节轨迹。实物实验证明六足机器人可以完成行走。【分层网络】【MPC】[111]

[2019 RAL] Fast and continuous foothold adaptation for dynamic locomotion through cnns*
为了实现在复杂地形下的鲁棒运动，magana等人使用CNN实现了基于视觉反馈的实时立足点调整。相较于传统启发式方法，其运算速度快，可以在运动中连续计算和校正立足点。该方案将建图得到的高度图导入CNN中，CNN输出调整的足端位置，用于生成关节轨迹。在实物机器人HyQ上进行了测试与验证。【视觉融合】[112]

[2019 RSS] Learning to Walk via Deep Reinforcement Learning*
为了利用RL无需模型及复杂的设计的特点，并规避训练对参数的敏感性。Haarnoja等人提出了一种基于最大熵RL的方法（SAC+自动熵调节）训练四足机器人控制器。Minitaur可以在一组超参数下的快速训练中实现稳定步态缺点：simple gait [113]

[2022 RSS] Rapid Locomotion via Reinforcement Learning*
Margolis使用关于速度的自适应课程和在线系统辨识技巧，通过深度强化学习实现了MIT mini cheetah的高速运动及在不同地面下的鲁棒运动。【课程学习】【系统辨识】【策略迁移】[114]

[2021 CoRL] Learning to Jump from Pixels*
为了在崎岖且不连续的地面上进行运动，需要视觉输入。margolis等人提出了基于深度的脉冲控制（DIC），这是一种合成高度敏捷的视觉引导运动行为的方法。策略网络生成动作输入到Whole-body Trajectory Generator中生成机身轨迹，再通过底层的MPC+WBIC控制器追踪生成的动态机身轨迹完成运动。DIC 提供了无模型学习的灵活性，但通过基于模型的地面反作用力的显式优化来规范行为。【视觉融合】【崎岖路面】[14]

[2018 ACM TG] DeepMimic: example-guided deep reinforcement learning of physics-based character skills*
为了模仿参考动作片段中的运动，peng等人使用强化学习方法设定模仿目标和任务目标相结合的奖励函数，实现机器人在动力学系统中与参考相似的运动。【模仿学习】[115]

[2021 TNNLS] Generic Neural Locomotion Control Framework for Legged Robots
Thor等人提出了一种通用的神经网络控制框架，可以用于各式足式机器人，包括六足机器人和四足机器人。这个工作提出了一种基于SO（2）振荡器和RBF函数的通用步态规划器框架和基于PIBB黑箱算法的自动参数学习方法，可以在少量的参数和简单的奖励函数下进行快速学习。但是，其只能够接受简单的反馈，难以对机身进行基于反馈的平衡控制，因此在四足机器人上效果较差。【CPG】[116]

[2018 ICRA] Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation*
针对具体的奖励函数难以制定的场合，模仿学习是一个有效的解决方案。然而很多时候状态-动作对无法获得以进行监督学习。Liu等人提出了从观测中进行模仿，即一种基于视频预测、上下文翻译和深度强化学习的模仿学习方法。实验证明了可以仅通过观察人类工具使用的视频来学习涉及工具使用的机器人技能。【模仿学习】[117]

[2022 arXiv] Imitate and Repurpose: Learning Reusable Robot Movement Skills From Human and Animal Behaviors*
为了利用来自人类和动物的先验信息来训练足式机器人的运动。Bohez等人利用动态捕捉得到的数

据，采用模仿学习的方式，实现了四足机器人和人形机器人的运动控制器设计。这种方法不需要大量的奖励函数设计就可以产生可以复用的策略。【模仿学习】[118]

[2023 arXiv] Learning Torque Control for Quadrupedal Locomotion*
传统RL输出参考轨迹，并使用下层PD控制器进行跟踪，这种低频率的策略难以实现高动态的运动，且PD参数难以设计。Chen等使用RL直接输出关节力矩，免去了下层的PD控制器。实验验证了机器人可以在不同地面下完成速度跟踪任务。

来自加州大学伯克利分校和美国加利福尼亚州的学者们根据基于模型控制的最新进展，通过引入基于扭矩的强化学习（RL）框架，探索基于位置的RL 范例的替代方案。其中RL 策略能够直接预测高频的关节扭矩，从而规避PD控制器的使用。他们所提出的学习扭矩控制框架已经通过大量实验得到了验证，其中四足动物能够在遵循用户指定的命令的同时穿越各种地形并抵抗外部干扰。此外，与学习位置控制相比，学习扭矩控制显示出更高的潜力，并且有外部干扰时更稳健。这也是对四足运动的端到端学习扭矩控制的首次模拟到真实的尝试。【输出力矩】[119]

[2022 arXiv] Visual CPG-RL: Learning Central Pattern Generators for Visually-Guided Quadruped Navigation*
基于之前的CPG-RL，bellegarda等人进一步融合了视觉感知信息，提出了Visual CPG-RL，具体而言，调节CPG参数的策略除了接受上层指令和机身状态之外，还接受由视觉传感器得到的地形高度测量数据。这些数据是由高度信息组成的点云图（实物部署时由深度相机获得），实验比较了MLP和LSTM组成的网络对最终运动效果的影响。添加了基于视觉的外部感知信息后，机器人可以实现避障和穿越崎岖地形。【融合视觉】【CPG】【Ijspeert】[120]

[2018 ACM TG] SFV: reinforcement learning of physical skills from videos*
为了解决动捕数据难以获取及利用大量的现有视频数据，peng等人提出了一种使物理模拟角色能够从视频（skill from video，SFV）中学习技能的方法。该方法利用深度网络估计视频中的对象位姿，并将其作为参考对象，通过基于模仿的强化学习实现了对给定运动的模仿。【模仿学习】[121]

[2022 arXiv] Imitation and Adaptation Based on Consistency: A Quadruped Robot Imitates Animals from Videos Using Deep Reinforcement Learning
为了利用自然界中动物运动的先验信息，yao等提出了一种视频模仿适应网络（video imitation adaptation network，VIAN），首先进行运动识别（recognition）将获取的视频片段输入CNN中，得到周期性或者非周期性的运动数据，再将数据导入Motion Adaptor中（x11）分解出周期性和非周期性数据，强化学习生成的策略用于生成附加补偿信号，奖励被设置为衡量模仿的相似程度。【模仿学习】[122]

[2020 SCI RO] Learning quadrupedal locomotion over challenging terrain*
为了在没有外界传感器的情况下实现四足机器人在复杂路面下的运动问题，lee等人使用了包括两步的teacher-student policy训练方法（私有观测），自动地形课程等方法实现了策略的零样本迁移并实现了机器人在复杂条件下的运动。【分步学习】【策略迁移】【地形课程】【teacher-student learning】[123]

[2022 SCI RO] Learning robust perceptive locomotion for quadrupedal robots in the wild*
为了利用外界的感知信息实现鲁棒的控制，miki等人提出了一种稳健且通用的基于RL的控制方案。该方案利用基于注意力的循环编码器，集成了机身传感器和感知信息。编码器经过端到端训练，并学会无缝组合不同的感知模式，而无需求助于启发式方法。其结果是一个具有高鲁棒性和速度的腿式运动控制器。【崎岖路面】【感知融合】[124]

[2021 ICRA] Dynamics Randomization Revisited: A Case Study for Quadrupedal Locomotion
在强化学习中（1）使用和不使用（2）什么条件下使用动态随机化方法训练出多种四足机器人步态控制策略，并将仿真策略应用到真实的四足机器人去探索使用动态随机化方法的必要性和可行性。通过实验发现即使不使用动态随机化的方法同样可以生成多种良好的运动控制策略并可以成功应用到真实的四足机器人平台，而使用动态随机性方法反而会对四足机器人运动性能产生一定影响，比如会降低步行的速度；同时通过实验发现在对某些参数使用动态随机化方法会生成良好的运动控制策略。最终得出结论以通过强化学习生成四足机器人运动控制策略为例，使用动态随机化方法有利有弊，需要具体分析造成由仿真策略转化到真实机器人策略失败的具体原因，从而去更好地选择最优的方法。[125]

[2021 CoRL] Fast and Efficient Locomotion via Learned Gait Transitions*

设计了一个分层学习框架，其使用强化学习训练高层的步态策略，低层使用模型预测控制器（MPC）优化电机转矩。他们以宇树A1机器人为硬件平台测试了该分层学习框架，在该测试中机器人的步态在爬行、对角小跑、飞行小跑步态中切换，并且机器人能将其速度提高至2.5m/s（约为5个身体长度/s）。【分层网络】【MPC】[16]

[2020 arXiv] Zero-shot terrain generalization for visual motion strategies
腿式机器人在非结构化地形上具有无与伦比的机动性，然而设计一个能在该地形上运动的机器人控制器是一个公认的巨大挑战。Alejandro Escontrela等研究员将这一挑战定义为多任务的强化学习问题，并将每个任务定义为机器人需要穿越的地形类型，同时文中还提出了一种新的程序地形生成算法，该算法能提供丰富的训练数据。在仿真环境中，研究员使用宇树Laikago四足机器人进行测试，测试表明所学习的控制器具有良好的零拍泛化能力，可以导航13个不同的目标环境，包括楼梯、崎岖的土地、杂乱的办公室、和室内空间与人类。下一步他们计划将这项研究推广到真实机器人的运动控制中。【策略迁移】【感知融合】[126]

[2023 CVPR] Neural Volumetric Memory for Visual Locomotion Control*
解决了学习运动控制器的问题，这些控制器可以推广到现实世界中常见的各种地形，并且在宇树科技的Laikago四足机器人运动平台和二次开发平台基础上提出了一种end-to-end 的视觉移动策略以实现高泛化性能表现。这种策略学习可以使腿式机器人能够在现实生活中的许多地形上进行导航，包括楼梯、崎岖的土地、障碍物、办公室等等，所有这些都不需要人工的视觉预输入。[53]

[2021 ICRA] Protective policy transfer
通过训练，优化任务奖励的任务策略，保护机器人形成不安全状态的策略，以及用于评估安全级别的安全评估模型机器人等三个模型提出了一种用于有意识地使控制策略适应新场景的转移学习算法，同时最大限度地减少了机器人的潜在损害.他们在DART仿真环境中使用unitree提供的A1模型，通过给每个电机引入功率限制，测试在该算法下仿真和训练的差异性。[127]

[2021 IROS] Animal gait using motion matching and model-based control on a quadruped robot*
结合了基于模型的控制器和计算机动画中常用的数据驱动技术提出了能为腿式机器人生产类似动物行走动作的控制方法。特别的是，该方法还能自动重新动物动作的关键特征，包括特定速度的步态、非周期性运动的无脚本脚步模态，以及整体身体运动的自然微小变化。ETH使用宇树Aliengo机器人在仿真器中进行实验。同时文中提出该方法能非常有效地实时运行，并且能很好地结合MPC-WBC控制方法。[128]

[2022 CoRL] Visual-Locomotion Learning to Walk on Complex Terrains with Vision*
视觉是有腿机器人安全有效地在不平坦地形（例如楼梯）上导航的基本感知方式之一。然而，训练机器人有效理解运动的高维视觉输入是一个具有挑战性的问题。在这项工作中，他们提出了一个框架来训练基于视觉的运动控制器，使四足机器人能够穿越不平坦的环境。关键思想是引入具有高级视觉策略和低级运动控制器的层次结构。高级视觉策略将感知到的视觉信号和机器人状态作为输入，并输出机器人所需的立足点和基本运动。然后这些由低级运动控制器实现，该控制器由摆动腿的位置控制器和基于MPC 的支撑腿扭矩控制器组成。使用深度强化学习训练视觉策略，并在各种不平坦的环境（例如随机放置的梅花桩、楼梯和移动平台）上展示此方法。同时还在真实的四足机器人上验证了此方法，以走过一系列间隙并爬上平台。[20]

[2023 ICRA] DribbleBot: Dynamic Legged Manipulation in the Wild
DribbleBot（用腿部机器人灵巧地操纵球）是一个腿部机器人系统，可以运球。来自美国麻省理工学院的学者们在模拟中使用强化学习并将其转移到现实世界中。他们克服了在不同的地形上核算可变的球运动动力学的关键挑战，以及如何在机载计算的限制下使用体载相机感知球。实验结果表明目前的四足平台非常适合研究涉及同时运动和操纵的动态控制问题，例如足球比赛。[129]

[2023 ICRA] Puppeteer and Marionette: Learning Anticipatory Quadrupedal Locomotion Based on Interactions of a Central Pattern Generator and Supraspinal Drive
shafiee等通过采用深度强化学习（DRL）训练了一种复制脊髓上驱动行为的神经网络策略。该策略可以调节CPG 动态，或直接更改驱动信号以绕过CPG 动态。他们的结果表明，对驱动信号的直接脊髓上贡献是高间隙穿越成功率的关键组成部分。然而，脊髓中的CPG 动力学有利于步态平稳性和能量效率。此外，他们的调查表明，感知前脚与间隙的距离是学习穿越间隙最重要和最充分的感官信息。研究结果支持以下生物学假设，即猫和马主要控制前腿来躲避障碍物，而后肢则根据前肢的信息进行内部记忆。他们的方法使四足机器人能够跨越长达20 厘米（身体长度的50%）的间隙，而无需任何显式动力学建模或模型预测控制（MPC）。【CPG】【Ijspeert】[130]

[2023 L4DC] Roll-Drop: accounting for observation noise with a single parameter
canmpanaro等人提出了Roll-Drop，一种在深度强化学习（DRL）中从模拟到真实的简单策略。DRL 是一种很有前途的控制机器人进行高动态和基于反馈的操作的方法，并且准确的模拟器对于提供丰富的数据以学习所需的行为至关重要。然而，模拟数据是无噪声的，并且通常显示出分布变化，这对传感器读数受噪声影响的真实机器上的部署提出了挑战。标准的解决方案是对后者建模并在训练期间注入。虽然这需要彻底的系统识别，但Roll-Drop 仅通过调整单个参数来增强对传感器噪声的鲁棒性。当在观察中注入高达25% 的噪声时，显示80% 的成功率，两次比基线具有更高的鲁棒性。将经过模拟训练的控制器部署在Unitree A1平台并在物理系统上评估这种改进的稳健性。[131]

[2022 arXiv] Sim-to-Real Transfer for Quadrupedal Locomotion via Terrain Transformer
Lai等提出了Terrain Transformer （TERT），这是一种用于四足运动的高容量变换器模型，可以控制各种地形。此外，为了更好地实现模拟到现实的转移，学者们提出了一个新颖的两阶段训练框架，包括一个离线预训练阶段和一个在线校正阶段。大量的模拟实验表明，TERT 在不同地形上的回报率优于最先进的基线，包括能耗和控制平滑度。在进一步的真实世界验证中，TERT 成功穿越了九个具有挑战性的地形，包括沙坑和楼梯，这都是在强基线无法完成的。【分步训练】【崎岖地形】[132]

[2023 arXiv] DeepTransition: Viability Leads to the Emergence of Gait Transitions in Learning Anticipatory Quadrupedal Locomotion Skills
Shafiee等通过利用深度强化学习和机器人工具，通过脊髓上驱动（大脑）、脊髓中的中央模式发生器、身体和外部感知之间的相互作用来研究四足动物步态转换的出现。与四足动物数据一致，学者们表明了四足机器人在平坦地形上的步行-小跑步态转换提高了生存能力和能源效率。此外，他们还研究了离散地形（即跨越连续的间隙）对施加步态转换的影响，并发现了小跑过渡的出现以避免非可行状态。与峰值力和能量效率等其他潜在标准相比，生存能力是平坦和离散间隙地形上步态转换后唯一改进的因素，这表明生存能力可能是步态转换的主要和普遍目标，而其他标准是次要目标以及生存能力的结果。学者们在从模拟到真实的硬件实验中部署了控制器，并在具有挑战性的场景中展示了最先进的四足动物敏捷性，其中Unitree A1 四足动物的自主转换步态以跨越多达30 个连续间隙（身体长度的83.3%），速度超过1.3 m/s。【Ijspeert】[133]

[2023 arXiv] CAJun: Continuous Adaptive Jumping using a Learned Centroidal Controller
来自华盛顿大学和谷歌Deepmind的学者们提出了CAJun，这是一种新颖的分层学习和控制框架，能够让腿式机器人以自适应跳跃距离连续跳跃。CAJun 由高级质心策略和低级腿部控制器组成。学者们使用强化学习（RL ）来训练质心策略，该策略指定腿部控制器的步态计时、基本速度和摆动脚位置。腿部控制器根据步态时序优化摆动腿和站立腿的电机命令，以跟踪摆动脚目标和基本速度命令。此外，他们还重新制定了腿部控制器中的姿态腿部优化器，以将策略训练速度加快一个数量级。这个系统实现了学习的多功能性，同时享有控制方法的鲁棒性，使其可以轻松转移到真正的机器人上。在单个GPU上训练20分钟后，CAJun可以在Go1 机器人上以自适应距离实现连续的长距离跳跃，模拟与真实的差距很小。此外，该机器人可以跳跃最大宽度为70厘米的间隙，比现有方法宽40%以上。[134]

[2019 ICRA] SpaceBok: A Dynamic Legged Robot for Space Exploration
Arm等人制作了一种用于行星探索的四足机器人SpaceBok,它使用并行的机构来实现运动，可以实现高速的运动和较高的跳跃高度。[135]

References

[1] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "RMA: Rapid motor adaptation for legged robots," in *Proceedings of Robotics: Science and Systems.*

[2] A. Kumar, Z. Li, J. Zeng, D. Pathak, K. Sreenath, and J. Malik, "Adapting rapid motor adaptation for bipedal robots," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1161–1168.

[3] Z. Fu, X. Cheng, and D. Pathak, "Deep whole-body control: Learning a unified policy for manipulation and locomotion," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, pp. 138–149. [Online]. Available: https://proceedings.mlr.press/v205/fu23a.html

[4] Z. Luo, E. Xiao, and P. Lu, "Ft-net: Learning failure recovery and fault-tolerant locomotion for quadruped robots," *IEEE Robotics and Automation Letters*, pp. 1–8, 2023.

[5] S. Kareer, N. Yokoyama, D. Batra, S. Ha, and J. Truong, "ViNL: Visual navigation and locomotion over obstacles," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2018–2024.

[6] J. Carius, F. Farshidian, and M. Hutter, "MPC-net: A first principles guided policy search," vol. 5, no. 2, pp. 2897–2904. [Online]. Available: https://ieeexplore.ieee.org/document/9001182/

[7] A. Reske, J. Carius, Y. Ma, F. Farshidian, and M. Hutter, "Imitation learning from MPC for quadrupedal multi-gait control," in *2021 IEEE International Conference on Robotics and Automation (ICRA).* IEEE, pp. 5014–5020. [Online]. Available: https://ieeexplore.ieee.org/document/9561444/

[8] Z. Xie, X. Da, B. Babich, A. Garg, and M. van de Panne, "GLiDE: Generalizable quadrupedal locomotion in diverse environments with a centroidal model," publisher: arXiv Version Number: 3. [Online]. Available: https://arxiv.org/abs/2104.09771

[9] T. Anne, J. Wilkinson, and Z. Li, "Meta-learning for fast adaptive locomotion with uncertainties in environments and robot dynamics," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 4568–4575.

[10] A. Shirwatkar, V. K. Kurva, D. Vinoda, A. Singh, A. Sagi, H. Lodha, B. G. Goswami, S. Sood, K. Nehete, and S. Kolathaya, "Force control for robust quadruped locomotion: A linear policy approach," in *2023 IEEE International Conference on Robotics and Automation (ICRA).* IEEE, pp. 5113–5119. [Online]. Available: https://ieeexplore.ieee.org/document/10161080/

[11] D. Kang, J. Cheng, M. Zamora, F. Zargarbashi, and S. Coros, "RL + model-based control: Using on-demand optimal control to learn versatile legged locomotion." [Online]. Available: http://arxiv.org/abs/2305.17842

[12] Q. Yao, J. Wang, D. Wang, S. Yang, H. Zhang, Y. Wang, and Z. Wu, "Hierarchical terrain-aware control for quadrupedal locomotion by combining deep reinforcement learning and optimal control," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE, pp. 4546–4551. [Online]. Available: https://ieeexplore.ieee.org/document/9636738/

[13] F. Yin, A. Tang, L. Xu, Y. Cao, Y. Zheng, Z. Zhang, and X. Chen, "Run like a dog: Learning based whole-body control framework for quadruped gait style transfer," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE, pp. 8508–8514. [Online]. Available: https://ieeexplore.ieee.org/document/9636805/

[14] G. B. Margolis, T. Chen, K. Paigwar, X. Fu, D. Kim, S. b. Kim, and P. Agrawal, "Learning to jump from pixels," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, pp. 1025–1034. [Online]. Available: https://proceedings.mlr.press/v164/margolis22a.html

[15] S. Gangapurwala, M. Geisert, R. Orsolino, M. Fallon, and I. Havoutis, "Real-time trajectory adaptation for quadrupedal locomotion using deep reinforcement learning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 5973–5979. [Online]. Available: https://ieeexplore.ieee.org/document/9561639/

[16] Y. Yang, T. Zhang, E. Coumans, J. Tan, and B. Boots, "Fast and efficient locomotion via learned gait transitions," in *5th Annual Conference on Robot Learning*, 2021. [Online]. Available: https://openreview.net/forum?id=vm8Hr9YJHZ1

[17] X. Da, Z. Xie, D. Hoeller, B. Boots, A. Anandkumar, Y. Zhu, B. Babich, and A. Garg, "Learning a contact-adaptive controller for robust, efficient legged locomotion," in *Proceedings of the 2020 Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, J. Kober, F. Ramos, and C. Tomlin, Eds., vol. 155. PMLR, pp. 883–894. [Online]. Available: https://proceedings.mlr.press/v155/da21a.html

[18] S. Gangapurwala, M. Geisert, R. Orsolino, M. Fallon, and I. Havoutis, "RLOC: Terrain-aware legged locomotion using reinforcement learning and optimal control," vol. 38, no. 5, pp. 2908–2927. [Online]. Available: https://ieeexplore.ieee.org/document/9779429/

[19] A. Pandala, R. T. Fawcett, U. Rosolia, A. D. Ames, and K. A. Hamed, "Robust predictive control for quadrupedal locomotion: Learning to close the gap between reduced- and full-order models," vol. 7, no. 3, pp. 6622–6629.

[20] W. Yu, D. Jain, A. Escontrela, A. Iscen, P. Xu, E. Coumans, S. Ha, J. Tan, and T. Zhang, "Visual-locomotion: Learning to walk on complex terrains with vision," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 08–11 Nov 2022, pp. 1291–1302. [Online]. Available: https://proceedings.mlr.press/v164/yu22a.html

[21] S. Xu, L. Zhu, and C. P. Ho, "Learning efficient and robust multi-modal quadruped locomotion: A hierarchical approach," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 4649–4655.

[22] "Learning locomotion skills using deeprl: Does the choice of action space matter?"

[23] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," in *Robotics: Science and Systems XIV*. Robotics: Science and Systems Foundation. [Online]. Available: http://www.roboticsproceedings.org/rss14/p10.pdf

[24] P. Eckert and A. J. Ijspeert, "Benchmarking agility for multilegged terrestrial robots," *IEEE Transactions on Robotics*, vol. 35, no. 2, pp. 529–535, 2019.

[25] G. Sartoretti, W. Paivine, Y. Shi, Y. Wu, and H. Choset, "Distributed learning of decentralized control policies for articulated mobile robots," *IEEE Transactions on Robotics*, vol. 35, no. 5, pp. 1109–1122, 2019.

[26] Y. Yang, K. Caluwaerts, A. Iscen, T. Zhang, J. Tan, and V. Sindhwani, "Data efficient reinforcement learning for legged robots," in *Proceedings of the Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, L. P. Kaelbling, D. Kragic, and K. Sugiura, Eds., vol. 100. PMLR, pp. 1–10. [Online]. Available: https://proceedings.mlr.press/v100/yang20a.html

[27] A. Singla, S. Bhattacharya, D. Dholakiya, S. Bhatnagar, A. Ghosal, B. Amrutur, and S. Kolathaya, "Realizing learned quadruped locomotion behaviors through kinematic motion primitives," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 7434–7440. [Online]. Available: https://ieeexplore.ieee.org/document/8794179/

[28] S. Choi and J. Kim, "Trajectory-based probabilistic policy gradient for learning locomotion behaviors," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 1–7.

[29] T. Li, H. Geyer, C. G. Atkeson, and A. Rai, "Using deep reinforcement learning to learn high-level policies on the ATRIAS biped," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 263–269.

[30] X. Cheng, A. Kumar, and D. Pathak, "Legs as manipulator: Pushing quadrupedal agility beyond locomotion," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5106–5112.

[31] A. S. Mastrogeorgiou, Y. S. Elbahrawy, A. Kecskemethy, and E. G. Papadopoulos, "Slope handling for quadruped robots using deep reinforcement learning and toe trajectory planning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3777–3782. [Online]. Available: https://ieeexplore.ieee.org/document/9341645/

[32] Y. Tang, J. Tan, and T. Harada, "Learning agile locomotion via adversarial training," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 6098–6105.

[33] X. Song, Y. Yang, K. Choromanski, K. Caluwaerts, W. Gao, C. Finn, and J. Tan, "Rapidly adaptable legged robots via evolutionary meta-learning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3769–3776.

[34] T. Matiisen, A. Oliver, T. Cohen, and J. Schulman, "Teacher–student curriculum learning," vol. 31, no. 9, pp. 3732–3740. [Online]. Available: https://ieeexplore.ieee.org/document/8827566/

[35] W. Yu, J. Tan, Y. Bai, E. Coumans, and S. Ha, "Learning fast adaptation with meta strategy optimization," vol. 5, no. 2, pp. 2950–2957. [Online]. Available: https://ieeexplore.ieee.org/document/9001157/

[36] S. Gangapurwala, A. Mitchell, and I. Havoutis, "Guided constrained policy optimization for dynamic quadrupedal robot locomotion," vol. 5, no. 2, pp. 3642–3649. [Online]. Available: https://ieeexplore.ieee.org/document/9028178/

[37] D. Jain, K. Caluwaerts, and A. Iscen, "From pixels to legs: Hierarchical learning of quadruped locomotion," in *Proceedings of the 2020 Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, J. Kober, F. Ramos, and C. Tomlin, Eds., vol. 155. PMLR, pp. 91–102. [Online]. Available: https://proceedings.mlr.press/v155/jain21a.html

[38] S. Ha, P. Xu, Z. Tan, S. Levine, and J. Tan, "Learning to walk in the real world with minimal human effort," in *Proceedings of the 2020 Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, J. Kober, F. Ramos, and C. Tomlin, Eds., vol. 155. PMLR, pp. 1110–1120. [Online]. Available: https://proceedings.mlr.press/v155/ha21c.html

[39] J. Siekmann, Y. Godse, A. Fern, and J. Hurst, "Sim-to-real learning of all common bipedal gaits via periodic reward composition," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 7309–7315. [Online]. Available: https://ieeexplore.ieee.org/document/9561814/

[40] Z. Li, X. Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for robust parameterized locomotion control of bipedal robots," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2811–2817. [Online]. Available: https://ieeexplore.ieee.org/document/9560769/

[41] F. Shi, T. Homberger, J. Lee, T. Miki, M. Zhao, F. Farshidian, K. Okada, M. Inaba, and M. Hutter, "Circus ANYmal: A quadruped learning dexterous manipulation with its limbs," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2316–2323.

[42] H. Zhang, J. Wang, Z. Wu, Y. Wang, and D. Wang, "Terrain-aware risk-assessment-network-aided deep reinforcement learning for quadrupedal locomotion in tough terrain," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 4538–4545. [Online]. Available: https://ieeexplore.ieee.org/document/9636519/

[43] G. A. Castillo, B. Weng, W. Zhang, and A. Hereid, "Robust feedback motion policy design using reinforcement learning on a 3d digit bipedal robot," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5136–5143.

[44] Z. Fu, A. Kumar, J. Malik, and D. Pathak, "Minimizing energy consumption leads to the emergence of gaits in legged robots," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 08–11 Nov 2022, pp. 928–937. [Online]. Available: https://proceedings.mlr.press/v164/fu22a.html

[45] W. Yu, D. Jain, A. Escontrela, A. Iscen, P. Xu, E. Coumans, S. Ha, J. Tan, and T. Zhang, "Visual-locomotion: Learning to walk on complex terrains with vision," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, pp. 1291–1302. [Online]. Available: https://proceedings.mlr.press/v164/yu22a.html

[46] G. B. Margolis and P. Agrawal, "Walk these ways: Tuning robot control for generalization with multiplicity of behavior," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, 14–18 Dec 2023, pp. 22–31. [Online]. Available: https://proceedings.mlr.press/v205/margolis23a.html

[47] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 08–11 Nov 2022, pp. 91–100. [Online]. Available: https://proceedings.mlr.press/v164/rudin22a.html

[48] D. Hoeller, N. Rudin, C. Choy, A. Anandkumar, and M. Hutter, "Neural scene representation for locomotion on structured terrain," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8667–8674, 2022.

[49] G. Ji, J. Mun, H. Kim, and J. Hwangbo, "Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion," vol. 7, no. 2, pp. 4630–4637. [Online]. Available: https://ieeexplore.ieee.org/document/9714001/

[50] L. Gan, J. W. Grizzle, R. M. Eustice, and M. Ghaffari, "Energy-based legged robots terrain traversability modeling via deep inverse reinforcement learning," vol. 7, no. 4, pp. 8807–8814. [Online]. Available: https://ieeexplore.ieee.org/document/9813568/

[51] Y. Shao, Y. Jin, X. Liu, W. He, H. Wang, and W. Yang, "Learning free gait transition for quadruped robots via phase-guided controller," vol. 7, no. 2, pp. 1230–1237.

[52] M. Sorokin, J. Tan, C. K. Liu, and S. Ha, "Learning to navigate sidewalks in outdoor environments," vol. 7, no. 2, pp. 3906–3913.

[53] R. Yang, G. Yang, and X. Wang, "Neural volumetric memory for visual locomotion control," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1430–1440.

[54] C. S. Imai, M. Zhang, Y. Zhang, M. Kierebinski, R. Yang, Y. Qin, and X. Wang, "Vision-guided quadrupedal locomotion in the wild with multi-modal delay randomization," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 5556–5563. [Online]. Available: https://ieeexplore.ieee.org/document/9981072/

[55] N. Rudin, D. Hoeller, M. Bjelonic, and M. Hutter, "Advanced skills by learning locomotion and local navigation end-to-end," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 2497–2503. [Online]. Available: https://ieeexplore.ieee.org/document/9981198/

[56] H. Li, W. Yu, T. Zhang, and P. M. Wensing, "Zero-shot retargeting of learned quadruped locomotion policies using hybrid kinodynamic model predictive control," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11 971–11 977.

[57] T.-Y. Yang, T. Zhang, L. Luu, S. Ha, J. Tan, and W. Yu, "Safe reinforcement learning for legged locomotion," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 2454–2461. [Online]. Available: https://ieeexplore.ieee.org/document/9982038/

[58] A. Escontrela, X. B. Peng, W. Yu, T. Zhang, A. Iscen, K. Goldberg, and P. Abbeel, "Adversarial motion priors make good substitutes for complex reward functions," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 25–32.

[59] Y. Ji, Z. Li, Y. Sun, X. B. Peng, S. Levine, G. Berseth, and K. Sreenath, "Hierarchical reinforcement learning for precise soccer shooting skills using a quadrupedal robot," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1479–1486. [Online]. Available: https://ieeexplore.ieee.org/document/9981984/

[60] X. Huang, Z. Li, Y. Xiang, Y. Ni, Y. Chi, Y. Li, L. Yang, X. B. Peng, and K. Sreenath, "Creating a dynamic quadrupedal robotic goalkeeper with reinforcement learning." [Online]. Available: http://arxiv.org/abs/2210.04435

[61] Z. Wang, A. Li, Y. Zheng, A. Xie, Z. Li, J. Wu, and Q. Zhu, "Efficient learning of robust quadruped bounding using pretrained neural networks," *IET Cyber-Systems and Robotics*, vol. 4, no. 4, p. 331–338, sep 2022. [Online]. Available: https://doi.org/10.1049/csy2.12062

[62] L. Smith, J. C. Kew, X. Bin Peng, S. Ha, J. Tan, and S. Levine, "Legged robots that keep on learning: Fine-tuning locomotion policies in the real world," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 1593–1599. [Online]. Available: https://ieeexplore.ieee.org/document/9812166/

[63] P. Wu, A. Escontrela, D. Hafner, P. Abbeel, and K. Goldberg, "DayDreamer: World models for physical robot learning," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, pp. 2226–2240. [Online]. Available: https://proceedings.mlr.press/v205/wu23c.html

[64] G. Feng, H. Zhang, Z. Li, X. B. Peng, B. Basireddy, L. Yue, Z. SONG, L. Yang, Y. Liu, K. Sreenath, and S. Levine, "GenLoco: Generalized locomotion controllers for quadrupedal robots," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, pp. 1893–1903. [Online]. Available: https://proceedings.mlr.press/v205/feng23a.html

[65] C. Li, M. Vlastelica, S. Blaes, J. Frey, F. Grimminger, and G. Martius, "Learning agile skills via adversarial imitation of rough partial demonstrations," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, 14–18 Dec 2023, pp. 342–352. [Online]. Available: https://proceedings.mlr.press/v205/li23b.html

[66] Y. Fuchioka, Z. Xie, and M. Van de Panne, "Opt-mimic: Imitation of optimized trajectories for dynamic quadruped behaviors," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5092–5098.

[67] E. Vollenweider, M. Bjelonic, V. Klemm, N. Rudin, J. Lee, and M. Hutter, "Advanced skills through multiple adversarial motion priors in reinforcement learning," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5120–5126.

[68] C. Li, S. Blaes, P. Kolev, M. Vlastelica, J. Frey, and G. Martius, "Versatile skill control via self-supervised adversarial imitation of unlabeled mixed motions," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 2944–2950.

[69] G. Christmann, Y.-S. Luo, J. H. Soeseno, and W.-C. Chen, "Expanding versatility of agile locomotion through policy transitions using latent state representation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 5134–5140. [Online]. Available: https://ieeexplore.ieee.org/document/10160776/

[70] M. Seo, R. Gupta, Y. Zhu, A. Skoutnev, L. Sentis, and Y. Zhu, "Learning to walk by steering: Perceptive quadrupedal locomotion in dynamic environments," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5099–5105.

[71] I. M. Aswin Nahrendra, B. Yu, and H. Myung, "DreamWaQ: Learning robust quadrupedal locomotion with implicit terrain imagination via deep reinforcement learning," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 5078–5084. [Online]. Available: https://ieeexplore.ieee.org/document/10161144/

[72] D. Crowley, J. Dao, H. Duan, K. Green, J. Hurst, and A. Fern, "Optimizing bipedal locomotion for the 100m dash with comparison to human running," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12 205–12 211.

[73] S. Gangapurwala, L. Campanaro, and I. Havoutis, "Learning low-frequency motion control for robust and dynamic robot locomotion," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5085–5091.

[74] S. Surana, B. Lim, and A. Cully, "Efficient learning of locomotion skills through the discovery of diverse environmental trajectory generator priors," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 134–12 141.

[75] L. M. Smith, J. C. Kew, T. Li, L. Luu, X. B. Peng, S. Ha, J. Tan, and S. Levine, "Learning and adapting agile locomotion skills by transferring experience," in *Robotics: Science and Systems XIX, Daegu, Republic of Korea, July 10-14, 2023*, K. E. Bekris, K. Hauser, S. L. Herbert, and J. Yu, Eds. [Online]. Available: https://doi.org/10.15607/RSS.2023.XIX.051

[76] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Robust and versatile bipedal jumping control through reinforcement learning," in *Robotics: Science and Systems XIX, Daegu, Republic of Korea, July 10-14, 2023*, K. E. Bekris, K. Hauser, S. L. Herbert, and J. Yu, Eds. [Online]. Available: https://doi.org/10.15607/RSS.2023.XIX.052

[77] J. Wu, G. Xin, C. Qi, and Y. Xue, "Learning robust and agile legged locomotion using adversarial motion priors," vol. 8, no. 8, pp. 4975–4982.

[78] F. Bjelonic, J. Lee, P. Arm, D. Sako, D. Tateo, J. Peters, and M. Hutter, "Learning-based design and control for quadrupedal robots with parallel-elastic actuators," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1611–1618, 2023.

[79] C. Yang, C. Pu, G. Xin, J. Zhang, and Z. Li, "Learning complex motor skills for legged robot fall recovery," *IEEE Robotics and Automation Letters*, vol. 8, no. 7, pp. 4307–4314, 2023.

[80] W. Yu, C. Yang, C. McGreavy, E. Triantafyllidis, G. Bellegarda, M. Shafiee, A. J. Ijspeert, and Z. Li, "Identifying important sensory feedback for learning locomotion skills," vol. 5, no. 8, pp. 919–932. [Online]. Available: https://www.nature.com/articles/s42256-023-00701-w

[81] C. Yang, C. Pu, T. Wei, C. Wang, and Z. Li, "Learning quadruped locomotion using bio-inspired neural networks with intrinsic rhythmicity." [Online]. Available: http://arxiv.org/abs/2305.07300

[82] F. Vezzi, J. Ding, A. Raffin, J. Kober, and C. Della Santina, "Two-stage learning of highly dynamic motions with rigid and articulated soft quadrupeds." [Online]. Available: http://arxiv.org/abs/2309.09682

[83] J. Wu, Y. Xue, and C. Qi, "Learning multiple gaits within latent space for quadruped robots." [Online]. Available: http://arxiv.org/abs/2308.03014

[84] A. Agarwal, A. Kumar, J. Malik, and D. Pathak, "Legged locomotion in challenging terrains using egocentric vision," in *Proceedings of The 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, K. Liu, D. Kulic, and J. Ichnowski, Eds., vol. 205. PMLR, 14–18 Dec 2023, pp. 403–415. [Online]. Available: https://proceedings.mlr.press/v205/agarwal23a.html

[85] Z. Zhuang, Z. Fu, J. Wang, C. Atkeson, S. Schwertfeger, C. Finn, and H. Zhao, "Robot parkour learning," in *Conference on Robot Learning (CoRL)*, 2023.

[86] X. Cheng, K. Shi, A. Agarwal, and D. Pathak, "Extreme parkour with legged robots," *arXiv preprint arXiv:2309.14341*, 2023.

[87] T. Miki, J. Lee, L. Wellhausen, and M. Hutter, "Learning to walk in confined spaces using 3d representation." [Online]. Available: http://arxiv.org/abs/2403.00187

[88] J. Shi, C. Bai, H. He, L. Han, D. Wang, B. Zhao, M. Zhao, X. Li, and X. Li, "Robust quadrupedal locomotion via risk-averse policy learning," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.

[89] L. Schneider, J. Frey, T. Miki, and M. Hutter, "Learning risk-aware quadrupedal locomotion using distributional reinforcement learning." [Online]. Available: http://arxiv.org/abs/2309.14246

[90] D. Youm, H. Jung, H. Kim, J. Hwangbo, H.-W. Park, and S. Ha, "Imitating and finetuning model predictive control for robust and symmetric quadrupedal locomotion," vol. 8, no. 11, pp. 7799–7806.

[91] Z. Luo, Y. Dong, X. Li, R. Huang, Z. Shu, E. Xiao, and P. Lu, "MorAL: Learning morphologically adaptive locomotion controller for quadrupedal robots on challenging terrains," pp. 1–8.

[92] A. Hartmann, D. Kang, F. Zargarbashi, M. Zamora, and S. Coros, "Deep compliant control for legged robots."

[93] G. Bellegarda and A. Ijspeert, "CPG-RL: Learning central pattern generators for quadruped locomotion," vol. 7, no. 4, pp. 12 547–12 554.

[94] M. Aractingi, P.-A. Léziart, T. Flayols, J. Perez, T. Silander, and P. Souères, "Controlling the solo12 quadruped robot with deep reinforcement learning," vol. 13, no. 1, p. 11945, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-023-38259-7

[95] Y. Jin, X. Liu, Y. Shao, H. Wang, and W. Yang, "High-speed quadrupedal locomotion by imitation-relaxation reinforcement learning," vol. 4, no. 12, pp. 1198–1208.

[96] H. Shi, B. Zhou, H. Zeng, F. Wang, Y. Dong, J. Li, K. Wang, H. Tian, and M. Q.-H. Meng, "Reinforcement learning with evolutionary trajectory generator: A general approach for quadrupedal locomotion," vol. 7, no. 2, pp. 3085–3092, number: 2. [Online]. Available: https://ieeexplore.ieee.org/document/9693519/

[97] V. Tsounis, M. Alge, J. Lee, F. Farshidian, and M. Hutter, "DeepGait: Planning and control of quadrupedal gaits using deep reinforcement learning," vol. 5, no. 2, pp. 3699–3706. [Online]. Available: https://ieeexplore.ieee.org/document/9028188/

[98] D. Jain, A. Iscen, and K. Caluwaerts, "Hierarchical reinforcement learning for quadruped locomotion," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 7551–7557. [Online]. Available: https://ieeexplore.ieee.org/document/8967913/

[99] P. Brakel, S. Bohez, L. Hasenclever, N. Heess, and K. Bousmalis, "Learning coordinated terrain-adaptive locomotion by imitating a centroidal dynamics planner," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 10 335–10 342. [Online]. Available: https://ieeexplore.ieee.org/document/9981648/

[100] G. Bellegarda, Y. Chen, Z. Liu, and Q. Nguyen, "Robust high-speed running for quadruped robots via deep reinforcement learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 10 364–10 370. [Online]. Available: https://ieeexplore.ieee.org/document/9982132/

[101] X. B. Peng, E. Coumans, T. Zhang, T.-W. E. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," in *Robotics: Science and Systems*.

[102] W. Tan, X. Fang, W. Zhang, R. Song, T. Chen, Y. Zheng, and Y. Li, "A hierarchical framework for quadruped locomotion based on reinforcement learning," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 8462–8468. [Online]. Available: https://ieeexplore.ieee.org/document/9636757/

[103] Z. Xie, H. Y. Ling, N. H. Kim, and M. Panne, "ALLSTEPS: Curriculum-driven learning of stepping stone skills," vol. 39, no. 8, pp. 213–224. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/cgf.14115

[104] N. Rudin, H. Kolvenbach, V. Tsounis, and M. Hutter, "Cat-like jumping and landing of legged robots in low gravity using deep reinforcement learning," vol. 38, no. 1, pp. 317–328. [Online]. Available: https://ieeexplore.ieee.org/document/9453856/

[105] J. Wang, C. Hu, and Y. Zhu, "CPG-based hierarchical locomotion control for modular quadrupedal robots using deep reinforcement learning," vol. 6, no. 4, pp. 7193–7200. [Online]. Available: https://ieeexplore.ieee.org/document/9465716/

[106] A. Iscen, G. Yu, A. Escontrela, D. Jain, J. Tan, and K. Caluwaerts, "Learning agile locomotion skills with a mentor," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 2019–2025.

[107] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," vol. 4, no. 26, p. eaau5872, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/10.1126/scirobotics.aau5872

[108] C. Yang, K. Yuan, Q. Zhu, W. Yu, and Z. Li, "Multi-expert learning of adaptive legged locomotion," vol. 5, no. 49, p. eabb2174, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/10.1126/scirobotics.abb2174

[109] S. Choi, G. Ji, J. Park, H. Kim, J. Mun, J. H. Lee, and J. Hwangbo, "Learning quadrupedal locomotion on deformable terrain," vol. 8, no. 74, p. eade2256. [Online]. Available: https://www.science.org/doi/10.1126/scirobotics.ade2256

[110] A. Iscen, K. Caluwaerts, J. Tan, T. Zhang, E. Coumans, V. Sindhwani, and V. Vanhoucke, "Policies modulating trajectory generators," in *Conference on Robot Learning*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:53110129

[111] T. Li, N. Lambert, R. Calandra, F. Meier, and A. Rai, "Learning generalizable locomotion skills with hierarchical reinforcement learning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 413–419. [Online]. Available: https://ieeexplore.ieee.org/document/9196642/

[112] O. A. V. Magaña, V. Barasuol, M. Camurri, L. Franceschi, M. Focchi, M. Pontil, D. G. Caldwell, and C. Semini, "Fast and continuous foothold adaptation for dynamic locomotion through cnns," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2140–2147, 2019.

[113] T. Haarnoja, S. Ha, A. Zhou, J. Tan, G. Tucker, and S. Levine, "Learning to walk via deep reinforcement learning," *arXiv preprint arXiv:1812.11103*, 2018.

[114] G. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, "Rapid locomotion via reinforcement learning," in *Robotics: Science and Systems*, 2022.

[115] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, "DeepMimic: example-guided deep reinforcement learning of physics-based character skills," vol. 37, no. 4, pp. 1–14, number: 4. [Online]. Available: https://dl.acm.org/doi/10.1145/3197517.3201311

[116] M. Thor, T. Kulvicius, and P. Manoonpong, "Generic neural locomotion control framework for legged robots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 4013–4025, 2021.

[117] Y. Liu, A. Gupta, P. Abbeel, and S. Levine, "Imitation from observation: Learning to imitate behaviors from raw video via context translation," 07 2017.

[118] S. Bohez, S. Tunyasuvunakool, P. Brakel, F. Sadeghi, L. Hasenclever, Y. Tassa, E. Parisotto, J. Humplik, T. Haarnoja, R. Hafner, M. Wulfmeier, M. Neunert, B. Moran, N. Siegel, A. Huber, F. Romano, N. Batchelor, F. Casarini, J. Merel, R. Hadsell, and N. Heess, "Imitate and repurpose: Learning reusable robot movement skills from human and animal behaviors," 2022.

[119] S. Chen, B. Zhang, M. W. Mueller, A. Rai, and K. Sreenath, "Learning torque control for quadrupedal locomotion," issue: arXiv:2203.05194. [Online]. Available: http://arxiv.org/abs/2203.05194

[120] G. Bellegarda and A. Ijspeert, "Visual cpg-rl: Learning central pattern generators for visually-guided quadruped navigation," 12 2022.

[121] X. B. Peng, A. Kanazawa, J. Malik, P. Abbeel, and S. Levine, "Sfv: Reinforcement learning of physical skills from videos," *ACM Trans. Graph.*, vol. 37, no. 6, Nov. 2018.

[122] Q. Yao, J. Wang, S. Yang, C. Wang, H. Zhang, Q. Zhang, and D. Wang, "Imitation and adaptation based on consistency: A quadruped robot imitates animals from videos using deep reinforcement learning," 2022.

[123] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," vol. 5, no. 47, p. eabc5986, number: 47. [Online]. Available: https://www.science.org/doi/10.1126/scirobotics.abc5986

[124] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," vol. 7, no. 62, p. eabk2822.

[125] Z. Xie, X. Da, M. van de Panne, B. Babich, and A. Garg, "Dynamics randomization revisited:a case study for quadrupedal locomotion," 2021.

[126] A. Escontrela, G. Yu, P. Xu, A. Iscen, and J. Tan, "Zero-shot terrain generalization for visual locomotion policies," 2020.

[127] W. Yu, C. K. Liu, and G. Turk, "Protective policy transfer," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 10 595–10 602.

[128] D. Kang, S. Zimmermann, and S. Coros, "Animal gaits on quadrupedal robots using motion matching and model-based control," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 8500–8507.

[129] Y. Ji, G. B. Margolis, and P. Agrawal, "Dribblebot: Dynamic legged manipulation in the wild," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5155–5162.

[130] M. Shafiee, G. Bellegarda, and A. Ijspeert, "Puppeteer and marionette: Learning anticipatory quadrupedal locomotion based on interactions of a central pattern generator and supraspinal drive," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 1112–1119.

[131] L. Campanaro, D. D. Martini, S. Gangapurwala, W. Merkt, and I. Havoutis, "Roll-drop: accounting for observation noise with a single parameter," in *Proceedings of The 5th Annual Learning for Dynamics and Control Conference*, ser. Proceedings of Machine Learning Research, N. Matni, M. Morari, and G. J. Pappas, Eds., vol. 211. PMLR, 15–16 Jun 2023, pp. 718–730. [Online]. Available: https://proceedings.mlr.press/v211/campanaro23a.html

[132] H. Lai, W. Zhang, X. He, C. Yu, Z. Tian, Y. Yu, and J. Wang, "Sim-to-real transfer for quadrupedal locomotion via terrain transformer," 2023.

[133] M. Shafiee, G. Bellegarda, and A. Ijspeert, "Deeptransition: Viability leads to the emergence of gait transitions in learning anticipatory quadrupedal locomotion skills," 2023.

[134] Y. Yang, G. Shi, X. Meng, W. Yu, T. Zhang, J. Tan, and B. Boots, "Cajun: Continuous adaptive jumping using a learned centroidal controller," *arXiv preprint arXiv:2306.09557*, 2023.

[135] P. Arm, R. Zenkl, P. Barton, L. Beglinger, A. Dietsche, L. Ferrazzini, E. Hampp, J. Hinder, C. Huber, D. Schaufelberger, F. Schmitt, B. Sun, B. Stolz, H. Kolvenbach, and M. Hutter, "Spacebok: A dynamic legged robot for space exploration," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6288–6294.