文章编号:1671-7554(2020)08-0061-06 DOI:10.6040/j.issn.1671-7554.0.2020.0607

张伟,山东大学教授、博士研究生导师,香港中文大学博士,美国加州大学伯克利分校博士后。主要从事模式识别、计算机视觉、机器学习、机器人等领域的研究,主持国家自然基金重大项目课题、联合基金重点项目、国家重点研发计划课题、山东省重大专项等 10 余项。在 IEEE TPAMI、TNNLS、TIP、TCYB、CVPR、ICCV、ECCV、IJCAI、AAAI、ICRA、IROS 等人工智能与机器人领域权威期刊和会议上发表论文 80 余篇,获美国、中国等发明专利授权 10 余项。曾获香港青年科学家提名奖、2 次 IEEE 最佳论文奖、国际学术竞赛冠军等。担任模式识别与机器智能专委会委员、计算机视觉专委会委员、成像探测与感知专委会委员,以及 PRL、Neurocomputing、《控制理论与应用》等期刊编委/特邀编委等。

基于深度强化学习的四足机器人运动控制发展现状与展望

张伟,谭文浩,李贻斌
(山东大学控制科学与工程学院,山东 济南 250061)

摘要: 受类脑计算启发的深度强化学习在人工智能、机器人等诸多领域中都取得了巨大的成功,该方法通过结合深度学习与强化学习获得了优异的场景感知能力与任务决策能力。本文首先介绍了两类应用较为广泛的深度强化学习方法及其基本原理,并通过回顾深度强化学习在四足机器人运动控制上的应用现状讨论了该方法的研究进展,最后通过总结现有方法及腿足机器人控制特点,对深度强化学习在四足机器人上的应用前景进行了展望。

关键词: 机器学习;深度强化学习;四足机器人;运动控制;步态学习
中图分类号: R574 **文献标志码:** A

Locmotion control of quadruped robot based on deep reinforcement learning: review and prospect

ZHANG Wei, TAN Wenhao, LI Yibin
(School of Control Science and Engineering, Shandong University, Jinan 250061, Shandong, China)

Abstract: Brain-inspired deep reinforcement learning has recently led to a wide range of successes in different domains such as artificial intelligence and robotics. The method combining both advantages of deep learning and reinforcement learning gets strong capability of perception and decision-making. In this paper, we first provide a brief overview of two

收稿日期: 2020-04-16; **网络出版时间:** 2020-08-06 14:03:28
网络出版地址: http://kns.cnki.net/kcms/detail/37.1390.R.20200805.0858.002.html
基金项目: 科技部重点研发计划(2017YFB1300205);山东省科技重大专项(新兴产业)(2018CXGC1503)
通讯作者: 张伟。E-mail:davidzhang@sdu.edu.cn

kinds of widely used deep reinforcement learning method and their fundamentals, then introduce the current status of deep reinforcement learning applying on quadruped robots. Finally, by summarizing the existing methods and the characteristics of locomotion for quadruped robots, we present future potential of deep reinforcement learning on quadruped robots.

Key words: Machine learning; Deep reinforcement learning; Quadruped robot; Locomotion control; Gait learning

机器学习是人工智能领域的核心方法之一,是通过对已有的知识和经验进行学习,不断提高自身性能的过程。近年来,受动物神经网络构造启发产生的人工神经网络[1]成为机器学习领域的热点研究方向之一,其中的深度学习方法已经在图像分析[2-3]、语音识别[4]及其他诸多领域[5-6]取得了许多令人瞩目的成绩。与此同时,强化学习作为机器学习领域的另一代表方法,展现出了其在决策问题中的巨大优势。深度强化学习方法将深度学习与强化学习相结合,其主要特点在于可以通过与环境交互,从无到有进行不断地尝试与学习。在四足机器人步态控制问题中,使用深度强化学习方法可以无需对机器人工作环境进行精确建模,促使机器人主动去适应新的环境,从而避免研究人员针对所有不同的场景编制程序。由于无需完全知晓环境与自身状态,四足机器人在与环境互动过程中通过不断完善自身知识库,理论上可以解决一切问题而无需专家干预。本文通过简要介绍近年来深度强化的发展历程及其在四足机器人步态控制问题上的研究进展,为相关研究人员提供参考。

1 深度强化学习发展历程

强化学习理论从上世纪开始就已有萌芽,发展至今的深度强化学习方法已应用于游戏[7-10]、机器人控制[11-14]以及视觉导航等领域[15-19]。

1.1 基于值函数的深度强化学习 由于计算能力与优化方法的限制,深度强化学习发展初期一直无法在四足机器人控制中取得满足实际应用需求的效果。直到2013年,Google DeepMind的Mnih等[7]将深度学习方法与传统强化学习方法(reinforce learning,RL)中的Q学习[20]方法相结合,提出了深度Q网络(deep Q network,DQN),用于基于视觉的Atari游戏测试,最终在多个游戏中取得了超越人类的学习效果。

在DQN训练过程中,采用了经验回放机制,通过与环境交互获得存储记忆 $e_t=(s_t,a_t,r_t,s_{t+1})$,其中,$s$ 代表观察到的状态(state),a 表示该状态下DQN输出的动作(action),r 表示该动作与环境交互获得的奖励值(reward),而 t 代表当前状态-行为对所处的时间步,所以 s_{t+1} 就表示在状态 s_t 时采取动作 a_t 后所抵达的下一状态。其中奖励获得规则由人根据任务目标提前设定。然后将存储记忆储存于记忆库 $D=\{e_1,e_2,e_3\cdots e_T\}$ 中,训练时从中随机抽取一批记忆样本进行训练,以降低训练数据之间的关联性。在通过深度卷积网络近似当前动作值函数 Q 的同时,采用一个结构相同的网络保存 N 个时间步之前的动作值函数 Q 的参数,称为目标 Q 网络,用 $Q(s',a'/\theta_i^-)$ 表示,其中 θ_i^- 表示在第 i 次迭代时的网络参数,$y_i=r+\gamma\max_{a'}Q(s',a'/\theta_i^-)$ 为目标值,通过降低损失函数 $(y_i-Q(s,a/\theta_i))^2$ 的期望值来优化当前网络参数。

在DQN的基础上,学者们先后提出了一系列基于值函数的深度学习方法,如Hasselt等[21]提出的深度双Q网络(deep double Q network,DDQN)、Haus-knecht等[22]提出的深度循环Q网络(deep recurrent Q network,DRQN)等。

基于值函数的强化学习方法学习高效,效果稳定,适用于游戏等离散动作空间的决策任务,但由于该方法需要对所有动作进行值函数估计,难以应对连续动作空间决策任务。

1.2 包含策略梯度的深度强化学习 除了基于值函数的深度强化学习方法,研究人员还提出了基于策略梯度[23]的深度强化学习方法,它通过计算当前策略奖励总期望对于策略参数 θ 的偏导数来更新策略,最终收敛于最优策略。理论上,由于基于策略梯度的方法可以直接优化提升奖励的总期望,应更适用于解决强化学习问题。对于能获得高奖励的行为策略,将提高选择该行为的概率,进而提高总的奖励期望。此外,基于策略梯度的强化学习方法无需根据所有动作的值函数选择决策行为,可以生成连续的动作,因而更适宜于机器人控制等连续动作空间的决策任务。

将基于值函数的深度强化学习方法与基于策略梯度的深度强化学习相结合的演员评论家(actor-critic,AC)算法,则可以发挥两者的优势,既可以汲取策略梯度的高效学习与适应连续动作空间的优点,又可以兼具基于值函数的方法高效稳定的特点。

但由于存在两个神经网络相互对抗,经验与记忆相似度高,AC 算法相对难以收敛。为解决这一问题,2016 年 Google DeepMind 的 Mnih 等[24]提出了异步优势演员评论家(asynchronous advantage actor-critic,A3C)方法,通过在多线程中分别进行环境交互与学习,进一步打乱记忆库中记忆顺序,降低记忆之间的关联性,提高了收敛性与收敛速度。

此外还有近端策略优化(proximal policy optimization, PPO)[25]、双延迟深度确定性策略梯度算法(twin delayed deep deterministic,TD3)[26]等强化学习算法,为四足机器人步态控制提供了良好的算法基础。

2 深度强化学习应用于四足机器人运动控制

传统腿足式机器人控制方法主要对机器人与环境进行分析与建模,然后通过正逆运动学进行姿态估计与轨迹规划[27-28]。但由于环境复杂多变,人工设计并不能覆盖所有情况,所以将深度强化学习方法应用于四足机器人是一种可行的手段。

2.1 机器人控制的深度强化学习应用特点 由于强化学习是在与环境交互过程中进行弱监督的学习,缺乏人工干预与调控,所以在机器人控制方面存在一些区别于传统控制手段的难点,例如学习困难、易陷入次优策略等。

为解决此类问题,2018 年,UC Berkeley 的 Peng 等[29]利用 PPO 的强化学习方法在仿真环境中训练多种模型,包括标准人形模型、猎豹模型以及 Atlas 机器人等,取得了惊人的效果。其机器人可以完成行走、奔跑、后空翻以及侧踢等行为,并在其他模型上习得了投掷物体、高空跳跃等行为,表现甚至超越人类。其主要贡献在于提出了参考状态初始化的重要性并验证了提前终止训练的重要性。

强化学习是使智能体从无到有开始学习,因此不局限于人类想象范围内的行为模式,所以在人们设定的奖励规则之外,智能体常常会学到一些超出人们设想的行为。在一些情况下这种行为是被鼓励的,有利于扩展解决问题的手段。但在很多情况下,这提高了研究人员的算法设计难度,因为一些任务的解决手段已提前有所设定,无需也不再鼓励智能体花费大量的时间与计算资源去进行策略空间的大范围探索,仅需在一个有所限定的动作与策略空间内进行探索与优化。譬如在设计学习四足机器人步态算法时,无需使机器人从零开始探索,在现有的步态方法的基础上进行改进与优化即可。所以 Peng 等提出在强化学习训练时,可以对智能体进行与参考动作相关的初始化,让智能体能够从现有策略入手,通过模仿现有策略中智能体关节的位置、速度等信息,学习已有的策略,同时加入新的任务目标,保证其在学习已有策略的同时能够适应复杂多变的环境。通过将复杂动作分解,以不同阶段的形式作为智能体初始化的条件,客观上也将复杂任务予以分解,降低任务的学习难度,更加符合人类的学习习惯。

而 Peng 等所提的提前终止也是为了提高学习效率,降低学习难度。由于强化学习要求智能体独立探索,所以难免与人类一样学到一些“坏习惯”或者陷入一些错误策略中难以脱出。在这种情况下,智能体提前终止训练可以降低智能体试错的成本,尽量避免学习失败的情况,从“好的”策略开始再次学习。此外,其他研究人员[30-31]也通过各种手段解决深度强化学习在应用于机器人时的各类难题。

2.2 步态控制的深度强化学习应用 由于强化学习模型为了学习一个新的任务目标需要与环境进行大量的互动,所以人们通常会让智能体在仿真环境中进行学习与试错,以此来避免在真实环境中学习时因智能体犯错而导致的损失。另外,也可以通过对仿真过程进行加速,用于节约真实环境中的训练成本。但是,即使最为出色的仿真平台也无法严格还原真实场景,无论是仿真的精度还是真实性,仿真平台都与真实环境有一定的差距,所以在仿真环境中可行的策略不一定能够在真实环境中应用。为解决这一问题,2018 年,谷歌的 Tan 等[32]提出了一种可以将在仿真环境中学习到的四足机器人控制策略直接应用于真实场景的方法。其主要思想在于尽可能地使仿真场景与真实场景相接近,通过不断降低仿真与真实的差距,最终实现四足机器人控制从仿真到真实的跨越。Tan 等详细测量了机器人质量、电机摩擦系数、延迟、环境摩擦系数及电池电压等诸多参数,并在训练中对这些参数施加噪声以提高算法的鲁棒性。此外,Tan 等还使用了一种更加可靠的电机执行器模型,以降低电机这一仿真与真实差距最大的干扰。实验结果表明,通过仿真中对真实的不断近似,可以在一定程度上解决四足机器人控制方法从仿真到真实“跨越”的问题。

另外一种思路是直接在仿真环境中学习真实环境可用的行走策略。2019 年,Hwangbo 等[30]提出了一种新的四足机器人控制系统,致力于解决机器人从仿真到真实的跨越问题以及提高强化学习方法

的学习速度。该工作的创新之一在于,它基于ANYmal机器人对目前无法在仿真环境中进行模拟仿真的电机进行了提前学习,也就是在机器人学习控制策略之前,先对电机的输入所对应的输出进行有监督学习,以一个单独的神经网络代表仿真中的执行器,在现实中利用真实机器人所配备的执行器采集监督学习所需的标签,并将其用于拟合该执行器的输入输出函数,这一措施有效解决了由于非线性、无精准模型的执行器在仿真环境中难以仿真导致无法将学到的策略部署于真实环境中的问题。与此同时,为缓解强化学习方法学习速度慢、效率低的问题,Hwangbo等[33]使用刚体接触解算器进行四足机器人的仿真与学习,大大提高了机器人的学习速度。

此外还有直接在算法层面设计出适合于机器人控制的强化学习算法的解决方案。由于强化学习算法本身要求智能体在弱监督情况下通过与环境进行交互,自主获取训练数据,因而对算法自身的探索能力提出了要求。换言之,有了更强的探索能力,就会有更多的学习经验,也就能够得出更好的学习策略。为使算法获得更强的探索能力,Haarnoja等[31]提出了柔性演员评论家(soft actor-critic,Soft AC)算法,致力于提高策略本身的信息熵,从而扩展机器人控制算法的探索空间,进而提升机器人控制性能。Haarnoja等将信息熵的概念引入到强化学习方法中,以信息熵来衡量强化学习策略的信息量,并有意识地增加策略的信息熵,当信息熵高时,说明行动策略信息量大,也就是策略的探索能力较强。但由于过强的探索能力会使策略无法专注于完成任务而是仅进行探索,Haarnoja等[34]又设计了改进Soft AC方法,可自动学习如何平衡探索能力与完成任务能力,使得算法的两种能力更为平衡,最终在真实情况下对四足机器人Ghost进行训练,并取得了良好的效果。

2.3 较为复杂任务的学习方法应用 随着四足机器人行进控制方法的日益多样化,利用强化学习方法对机器人进行高级控制的需求也逐渐产生,人们不再满足于仅仅控制四足机器人简单移动,而是希望可以控制机器人走出某种特定的轨迹,或是完成一些特定的任务,这就产生了机器人控制的层次问题。执行复杂策略的高层理解认知决策行为与底层的行为产生、平衡控制并不一定需要进行端到端的处理。同时由于强化学习本身收敛困难,对于复杂任务,要求从零开始学习的可行性较低,因而对学习所需的网络结构与学习过程进行分层处理就是一种非常直观的选择。2019年,谷歌大脑的Jain等[35]对四足机器人控制提出了一种分层强化学习的解决方案,用于使Ghost机器人沿特定轨迹前进,利用高级策略网络控制低级策略网络,使用一个新的神经网络控制策略调制轨迹生成器[36],进而实现机器人控制。因此,该方法可以使高级策略网络将注意力集中于判断机器人行进方向,而无需学习沿该方向前进所需的电机控制指令。也就是说,当完成一种路径的学习后,底层控制网络无需改变就可以去学习其他路径,进而大幅降低学习新任务所需的时间。

在强化学习的基础上,为使强化学习可以更快地学习新任务,研究人员将元学习与强化学习结合,提高了任务学习过程中的样本利用率与训练速度。比如,为提升腿足机器人运动控制的学习效率,Frans等[37]使用高级策略网络选择需要执行的底层策略,利用多种底层策略的组合完成任务。通过在训练时进行“热身”,即在训练完整网络前对高级策略网络进行预训练,以达成优化训练效果、提升训练速度的目的。

3 深度强化学习在四足机器人上的应用前景

四足机器人传统控制方法目前已能适应多种工作环境[38-41],但对于每种环境都需要研究人员进行反复设计与调试。尽管传统控制方法可以满足小范围环境中机器人运动控制的任务需求,但不足以在更具挑战性的大范围真实环境中获取同样的控制效果。而基于学习方法的控制策略可以根据学习自动生成,无需机器人或工作环境的先验知识。理论上,相同的学习方法应能使机器人在每种环境中都学习出最优策略[42]。因此基于学习的机器人控制方法必然会在四足机器人控制中得到广泛应用。

虽然已有很多方法实现了四足机器人步态规划学习从计算机仿真向真实世界迁移的跨越,这种跨越仍然是一个理论方法与工程部署交织衔接的难题。尽管如此,作为理论到实践不可或缺的一步,这一难题也在逐步被研究人员攻克。随着理论方法的完善与实际部署的提升,更适应真实环境的机器人运动控制算法与更贴合实际的四足机器人模型都将不断产生,仿真到真实的距离也将不断缩短。

此外,虽然近年来基于学习的腿足机器人控制方法发展迅速,但传统控制手段依然值得借鉴。因此,将学习方法与传统方法的优势进行结合,利用基于学习的方法适应复杂多变的场景,同时继承传统

控制方法的稳定性,能够大幅提高四足机器人的运动控制能力。

综上,随着机器学习与类脑计算的不断发展,基于学习的四足机器人运动控制有望实现从仿真到真实的跨越、从单一方法到与传统方法相结合的跨越,未来将具有巨大发展潜力。

参考文献:

[1] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities [J]. NAS, 1982, 79(8): 2554-2558.

[2] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks [C] // Pereira F, Burges CJC, Bottou L. Advances in neural information processing systems. Lake Tahoe: Neural Information Processing Systems Conference, 2012: 1097-1105.

[3] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge [J]. IJCV, 2015, 115(3): 211-252.

[4] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks [C] // Krishnamurthy V, Plataniotis K. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver: IEEE, 2013: 6645-6649.

[5] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint arXiv: 1406.1078, 2014.

[6] Karpathy A, Toderici G, Shetty S, et al. Large-scale video classification with convolutional neural networks [C] // Basri R, Fermuller C, Martinez A. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1725-1732.

[7] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning [J]. arXiv preprint arXiv: 1312.5602, 2013.

[8] Wu Y, Zhang W, Song K. Master-slave curriculum design for reinforcement learning [C] // Lang J. IJCAI. Stockholm: International Joint Conferences on Artificial Intelligence, 2018: 1523-1529.

[9] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529-533.

[10] Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search [J]. Nature, 2016, 529(7587): 484-489.

[11] Lillicrap TP, Hunt JJ, Pritzel A, et al. Continuous control with deep reinforcement learning [J]. arXiv preprint arXiv: 1509.02971, 2015.

[12] Zhang J, Zhang W, Song R, et al. Grasp for stacking via deep reinforcement learning [C] // Kallio P, Burdet E. 2020 IEEE The International Conference on Robotics and Automation (ICRA), Virtual Conference: IEEE, 2020.

[13] Duan Y, Chen X, Houthooft R, et al. Benchmarking deep reinforcement learning for continuous control [C]. // Balcan N, Weinberger K. International Conference on Machine Learning. New York City: PMLR, 2016: 1329-1338.

[14] Gu S, Lillicrap T, Sutskever I, et al. Continuous deep Q-learning with model-based acceleration [C] // Balcan N, Weinberger K. International Conference on Machine Learning. New York City: PMLR, 2016: 2829-2838.

[15] Hansen S. Using deep Q-learning to control optimization hyperparameters [J]. arXiv preprint arXiv: 1602.04062, 2016.

[16] Wu Y, Rao Z, Zhang W, et al. Exploring the task cooperation in multi-goal visual navigation [C] // Hentenryck PV, Zhou ZH. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Hawaii: AAAI Press, 2019: 609-615.

[17] Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent [C] // Lee DD, Sugiyama M, Luxburg UV. Advances in neural information processing systems. Barcelona: Neural Information Processing Systems Conference, 2016: 3981-3989.

[18] Oh J, Guo X, Lee H, et al. Action-conditional video prediction using deep networks in atari games [C] // Cortes C, Lawrence ND, Lee DD. Advances in neural information processing systems. Montréal: Neural Information Processing Systems Conference, 2015: 2863-2871.

[19] Caicedo JC, Lazebnik S. Active object localization with deep reinforcement learning [C] // Ikeuchi K, Schnörr C, Sivic J. Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 2488-2496.

[20] Watkins CJCH, Dayan P. Q-learning [J]. Machine Learning, 1992, 8(3-4): 279-292.

[21] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning [C] // Schuurmans D, Wellman M. Thirtieth AAAI conference on Artificial Intelligence. Phoenix: AAAI Press, 2016.

[22] Hausknecht M, Stone P. Deep recurrent Q-learning for partially observable mdps [C] // Bonet B, Koenig S. 2015 AAAI Fall Symposium Series. Austin: AAAI Press, 2015.

[23] Sutton RS, McAllester DA, Singh SP, et al. Policy gradient methods for reinforcement learning with function approximation [C] // Leen TK, Dietterich TG, Tresp V. Advances in neural information processing systems. Denver: Neural Information Processing Systems Conference, 2000: 1057-1063.

[24] Mnih V, Badia AP, Mirza M, et al. Asynchronous methods for deep reinforcement learning [C] // Balcan N, Weinberger K. International Conference on Machine Learning. New York City: PMLR, 2016: 1928-1937.

[25] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms [J]. arXiv preprint arXiv: 1707.06347, 2017.

[26] Fujimoto S, Van Hoof H, Meger D. Addressing function approximation error in actor-critic methods [J]. arXiv preprint arXiv: 1802.09477, 2018.

[27] Jenelten F, Hwangbo J, Tresoldi F, et al. Dynamic locomotion on slippery ground [J]. IEEE Robotics and Automation Letters, 2019, 4(4): 4170-4176.

[28] Jin B, Sun C, Zhang A, et al. Joint torque estimation toward dynamic and compliant control for gear-driven torque sensorless quadruped robot [C] // Arai F. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). MACAO: IEEE, 2019: 4630-4637.

[29] Peng XB, Abbeel P, Levine S, et al. Deepmimic: example-guided deep reinforcement learning of physics-based character skills [J]. ACM Transactions on Graphics (TOG), 2018, 37(4): 1-14.

[30] Hwangbo J, Lee J, Dosovitskiy A, et al. Learning agile and dynamic motor skills for legged robots [J]. Science Robotics, 2019, 4(26): eaau5872.

[31] Haarnoja T, Zhou A, Abbeel P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor [J]. arXiv preprint arXiv: 1801.01290, 2018.

[32] Tan J, Zhang T, Coumans E, et al. Sim-to-real: learning agile locomotion for quadruped robots [J]. arXiv preprint arXiv: 1804.10332, 2018.

[33] Hwangbo J, Lee J, Hutter M. Per-contact iteration method for solving contact dynamics [J]. IEEE Robotics and Automation Letters, 2018, 3(2): 895-902.

[34] Haarnoja T, Zhou A, Hartikainen K, et al. Soft actor-critic algorithms and applications [J]. arXiv preprint arXiv: 1812.05905, 2018.

[35] Jain D, Iscen A, Caluwaerts K. Hierarchical reinforcement learning for quadruped locomotion [J]. arXiv preprint arXiv: 1905.08926, 2019.

[36] Iscen A, Caluwaerts K, Tan J, et al. Policies modulating trajectory generators [J]. arXiv preprint arXiv: 1910.02812, 2019.

[37] Frans K, Ho J, Chen X, et al. Meta learning shared hierarchies [J]. arXiv preprint arXiv: 1710.09767, 2017.

[38] Kolvenbach H, Hampp E, Barton P, et al. Towards jumping locomotion for quadruped robots on the moon [C] // Arai F. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). MACAO: IEEE, 2019: 5459-5466.

[39] Saputra AA, Toda Y, Takesue N, et al. A novel capabilities of quadruped robot moving through vertical ladder without handrail support [C] // Arai F. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). MACAO: IEEE, 2019: 1448-1453.

[40] Lee YH, Lee YH, Lee H, et al. Whole-body motion and landing force control for quadrupedal stair climbing [C] // Arai F. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). MACAO: IEEE, 2019: 4746-4751.

[41] Jenelten F, Hwangbo J, Tresoldi F, et al. Dynamic locomotion on slippery ground [J]. IEEE Robotics and Automation Letters, 2019, 4(4): 4170-4176.

[42] Ha S, Xu P, Tan Z, et al. Learning to walk in the real world with minimal human effort [J]. arXiv preprint arXiv: 2002.08550, 2020.

(编辑: 相峰)