Learning Skateboarding for Humanoid Robots through Massively Parallel Reinforcement Learning

William Thibault[1], Vidyasagar Rajendran[2], William Melek[1], Katja Mombaur[2,3]

Abstract— Learning-based methods have proven useful at generating complex motions for robots, including humanoids. Reinforcement learning (RL) has been used to learn locomotion policies, some of which leverage a periodic reward formulation. This work extends the periodic reward formulation of locomotion to skateboarding for the REEM-C robot. Brax/MJX is used to implement the RL problem to achieve fast training. Initial results in simulation are presented with hardware experiments in progress.

I. INTRODUCTION

Humanoid robots have been gaining popularity recently due to the advantage of having general-purpose robots that can operate in human-like environments to fill in labor shortages for undesirable tasks. For any application, the core skills required to interact and move about the environment involve manipulation and locomotion. Advances in machine learning have lead to the successful use of learning-based techniques, including reinforcement learning (RL), for learning these core skills. Many works have shown that bipedal walking can be learned for robust and stable walking on real robots making learning-based approaches appealing for generating complex motions for humanoids [1]–[3]. Due to the periodic nature of walking, periodic reward formulations, as originally suggested by Siekmann et al. [4], have been a popular way to learn bipedal walking as it allows for easily describing contact and flight phases of walking to produce a symmetrical gait. This periodic reward formulation proved crucial for humanoid robots that have large, heavy limbs like HRP-5P from Singh et al. 2022, 2023 [3], [5]. Recently, a periodic reward formulation that leveraged the Brax RL algorithms [6] with MJX [7], the XLA version of the MuJoCo simulator [8], was used to train the REEM-C humanoid robot to perform joystick walking in under 1 hour. Similarly to Issac Gym which has proven to be powerful for reinforcement learning problems like quadruped walking [9], Brax and MJX allow for massively parallel training with some benchmarks suggesting better training performance [10]. Periodic reward formulations allow for flexible description of repetitive gait cycles, leading to the
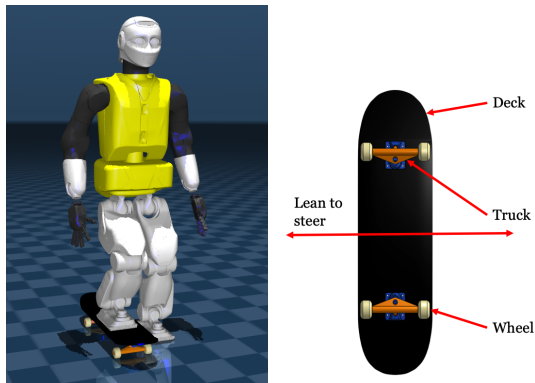

Fig. 1: REEM-C on skateboard (left), skateboard (right).

possibility of other periodic gait motions like skipping or hopping [4].

Skateboarding is a complex bipedal task that involves using one foot to push on the ground while the other foot is on the skateboard to propel forward, while also balancing side to side to maintain a straight trajectory. Previously, the small hybrid bipedal, drone robot LEONARDO was shown to skateboard by using its propellers to push itself rather than using its feet while leaning side to side to turn [11]. Another work demonstrated a skating like motion with HRP-2 using a skateboard-like device that does not involve lateral balance (a board with fixed wheels) and a kick board that has a handle [12]. However, skateboarding for a full size humanoid robot on a real skateboard has yet to be achieved. This work explores the possibility of leveraging RL with a periodic reward formulation for the cyclic motion of skateboarding on the REEM-C robot as an extension of the walking formulation of Thibault el. al [13].

II. METHOD

A. REEM-C Robot and Simulation Environment

The goal of learning joystick-style velocity skateboarding involves the actuated REEM-C robot and a passive skateboard. The REEM-C robot is a full-size humanoid robot with 30 degrees of freedom (DoF): two 6 DOF legs, two 7 DoF arms, a 2 DoF torso and a 2 DoF head. Similarly to Thibault et al. 2024 [13], only the 12 DoF for the legs are actuated with the other joints held fixed to simplify the learning problem. The initial position of the REEM-C on the skateboard and the model of the skateboard can be seen in Figure 1 where the right foot is on the deck of the skateboard and the left is on the ground. The skateboard has four passive wheels that are attached to the board, known as the deck, by two passive trucks. The trucks are used to hold the wheels while attached to the board so that when leaning more towards one

[1]William Thibault and William Melek are with Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada {wcthibau, wmelek}@uwaterloo.ca

[2]Vidyasagar Rajendran and Katja Mombaur are with the Department of Systems Design Engineering, CERC Human-Centred Robotics and Machine Intelligence, University of Waterloo, Canada {vrajendr, kmombaur}@uwaterloo.ca

[3]Katja Mombaur is also with the Institute for Anthropomatics and Robotics (IAR), Optimization and Biomechanics for Human-Centred Robotics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany katja.mombaur@kit.edu

edge of the deck, the trucks flex to turn the wheels and in turn the board in the direction of the leaning. Both the REEM-C robot and skateboard are modelled with complete dynamics with floating bases in the MJX simulator. The contacts of the feet with the deck and ground are unilateral as are the contacts of the wheels with the ground. For theses contacts, increased friction is applied for the contact of the right foot with the deck to simulate the effect of the grip tape that is normally on the top surface of the deck. With respect to the trucks, a stiffness was added to the joint to replicate the spring-like effect of trucks to remain centered when there is little leaning, but turned when leaning.

B. RL Problem Formulation

This problem formulation is a direct extension of Thibault et al. [13], thereby leveraging the same action space, observation space and rewards with a few exceptions. With a periodic reward formulation, contact and flight phase rewards are used to describe the behaviour of the feet for a cyclic motion. In this case, a single foot repeated pushing motion is formulated where the right foot remains on the skateboard and the left foot pushes on the ground then lifts repeatedly. For this pushing motion the command velocity that the robot's center of mass and the skateboard deck will attempt to track is limited to the forward direction only: $(x, y, \psi) = ([0.0, 1.0]$m/s, 0.0 m/s, 0.0 rad/s). The observation space is augmented with information about the skateboard including the 3 dimensional deck position and the 6 dimensional deck linear and angular velocity. Given the periodic reward formulation, the indicator function for gait phases is defined with a 0.75 s double support phase and a 1.0 s single support phase with the left foot in swing and the right foot on the skateboard. The exact same reward formulation and weights are sufficient to learn the skateboarding problem; however, the following rewards are added for faster learning along with improved behaviour for the deck's motion and the right foot to deck contact:

- Deck linear velocity tracking reward: This reward tracks the target linear velocity for the deck.
  $r_1 = e^{-||v_{xy,input} - v_{xy,deck}||_2^2/\sigma}$
  where $v_{xy,input}$ is the XY command velocity, $v_{xy,deck}$ is the XY deck velocity and σ is a scaling factor.
- Deck angular velocity tracking reward: This reward tracks the target angular velocity for the deck.
  $r_2 = e^{-(\omega_{z,input} - \omega_{z,deck})^2/\sigma}$
  where $\omega_{z,input}$ is the command angular velocity, $\omega_{z,deck}$ is the deck velocity and σ is a scaling factor.
- Deck foot world velocity reward: This is an additional reward for the right foot to move while in contact.
  $r_3 = ||v_{right}||_2^2$
  where v_{right} is the velocity of the right foot.
- Foot slip with respect to deck penalty: This reward penalizes translation of the right foot on the deck to keep the foot centered on the deck.
  $r_4 = ||v_{xy,deck} - v_{xy,right}||_2^2$
  where $v_{xy,deck}$ is the XY velocity of the deck and $v_{xy,right}$ is the XY velocity of the right foot.

- Foot rotation with respect to deck penalty: This reward penalizes roll and pitch velocities of the right foot on the deck to keep the foot contact flat.
  $r_5 = ||\omega_{xy,deck} - \omega_{xy,right}||_2^2$
  where $\omega_{xy,deck}$ is the roll and pitch angular velocities of the deck and $\omega_{xy,right}$ is the roll and pitch angular velocities of the right foot.

III. RESULTS

The RL training was set up with Brax and MJX for highly parallel training with 8192 parallel environments for 200,000,000 training steps using PPO similarly to Thibault et al. [13]. This RL training pipeline was selected as it allows for fast wall clock training times, while leveraging the simulation behaviour of MuJoCo. A skateboarding motion sequence can be seen in Figure 2.
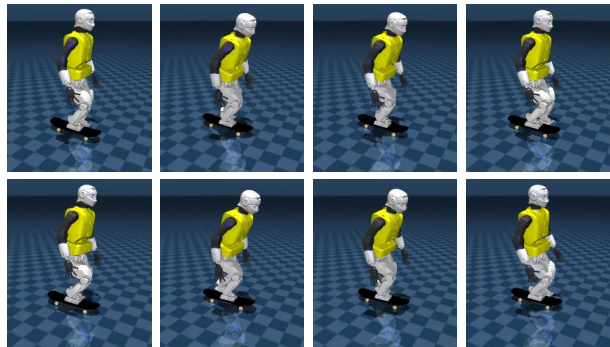
Fig. 2: REEM-C skateboarding forward at 0.4 m/s

In the simulation results, it can be seen that the left foot is used to push along the ground until the foot nears the back wheels of the skateboard then it is lifted and placed near the front wheels of the skateboard to perform the next push. The right foot remains flat and centered on the deck during the pushing motion. An interesting emergent behaviour occurs with the unactuated upper body during the push phase of the motion. The upper body leans forward to maintain balance while the robot pushes itself forward similar to the way a human would skateboard. The overall motion achieved is smooth and balanced with little turning.

IV. CONCLUSION

In this work, initial results for a humanoid skateboarding RL policy based on a periodic reward formulation was shown for the REEM-C robot. This work extends Thibault et al. [13] through a new gait cycle and added rewards for improved skateboarding performance while leveraging the fast, massively parallel training approach with Brax/MJX. On going work involves transferring the learned skateboarding abilities to the real REEM-C robot. Additionally, the skateboarding abilities are being extend to include a glide phase with both feet on the skateboard and the ability to turn.

V. ACKNOWLEDGEMENT

REFERENCES

[1] D. Rodriguez and S. Behnke, "Deepwalk: Omnidirectional bipedal gait by deep reinforcement learning," no. arXiv:2106.00534, June 2021, arXiv:2106.00534 [cs]. [Online]. Available: http://arxiv.org/abs/2106.00534

[2] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Learning humanoid locomotion with transformers," no. arXiv:2303.03381, Mar. 2023, arXiv:2303.03381 [cs]. [Online]. Available: http://arxiv.org/abs/2303.03381

[3] R. P. Singh, Z. Xie, P. Gergondet, and F. Kanehiro, "Learning bipedal walking for humanoids with current feedback," *IEEE Access*, vol. 11, p. 82013–82023, 2023, arXiv:2303.03724 [cs].

[4] J. Siekmann, Y. Godse, A. Fern, and J. Hurst, "Sim-to-real learning of all common bipedal gaits via periodic reward composition," no. arXiv:2011.01387, Mar. 2021, arXiv:2011.01387 [cs]. [Online]. Available: http://arxiv.org/abs/2011.01387

[5] R. P. Singh, M. Benallegue, M. Morisawa, R. Cisneros, and F. Kanehiro, "Learning bipedal walking on planned footsteps for humanoid robots," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. Ginowan, Japan: IEEE, Nov. 2022, p. 686–693. [Online]. Available: https://ieeexplore.ieee.org/document/10000067/

[6] C. D. Freeman, E. Frey, A. Raichuk, S. Girgin, I. Mordatch, and O. Bachem, "Brax - a differentiable physics engine for large scale rigid body simulation," 2021. [Online]. Available: http://github.com/google/brax

[7] MuJoCo XLA (MJX). Accessed: 2024-07-06. [Online]. Available: https://mujoco.readthedocs.io/en/stable/mjx.html

[8] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5026–5033.

[9] 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*. PMLR, Jan. 2022, p. 91–100. [Online]. Available: https://proceedings.mlr.press/v164/rudin22a.html

[10] K. Caluwaerts, A. Iscen, J. C. Kew, W. Yu, T. Zhang, D. Freeman, K.-H. Lee, L. Lee, S. Saliceti, V. Zhuang, N. Batchelor, S. Bohez, F. Casarini, J. E. Chen, O. Cortes, E. Coumans, A. Dostmohamed, G. Dulac-Arnold, A. Escontrela, E. Frey, R. Hafner, D. Jain, B. Jyenis, Y. Kuang, E. Lee, L. Luu, O. Nachum, K. Oslund, J. Powell, D. Reyes, F. Romano, F. Sadeghi, R. Sloat, B. Tabanpour, D. Zheng, M. Neunert, R. Hadsell, N. Heess, F. Nori, J. Seto, C. Parada, V. Sindhwani, V. Vanhoucke, and J. Tan, "Barkour: Benchmarking animal-level agility with quadruped robots," no. arXiv:2305.14654, May 2023, arXiv:2305.14654 [cs]. [Online]. Available: http://arxiv.org/abs/2305.14654

[11] K. Kim, P. Spieler, E.-S. Lupu, A. Ramezani, and S.-J. Chung, "A bipedal walking robot that can fly, slackline, and skateboard," *Science Robotics*, vol. 6, no. 59, p. eabf8136, Oct. 2021.

[12] N. Takasugi, K. Kojima, S. Nozawa, Y. Kakiuchi, K. Okada, and M. Inaba, "Real-time skating motion control of humanoid robots for acceleration and balancing," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon, South Korea: IEEE, Oct. 2016, p. 1356–1363. [Online]. Available: http://ieeexplore.ieee.org/document/7759223/

[13] W. Thibault, W. Melek, and K. Mombaur, "Learning Velocity-based Humanoid Locomotion: Massively Parallel Learning with Brax and MJX," 2024. [Online]. Available: https://arxiv.org/abs/2407.05148