Sim-to-Real Learning of All Common Bipedal Gaits via Periodic Reward Composition

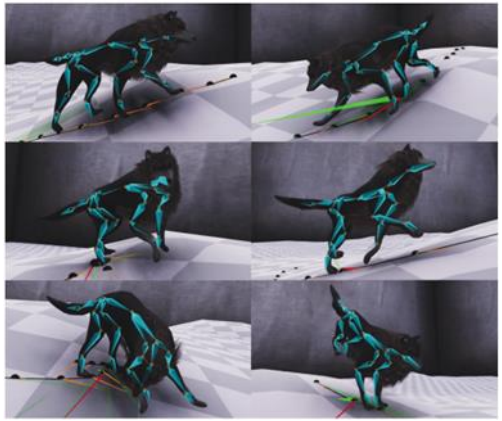
Jonah Siekmann*, Yesh Godse*, Alan Fern, Jonathan Hurst
Collaborative Robotics and Intelligent Systems Institute
Oregon State University

Article Presentation

Group 6 :
Xiaoyu Yan 395091
Zekun Wang 393910
Yuansheng Zhou 393916

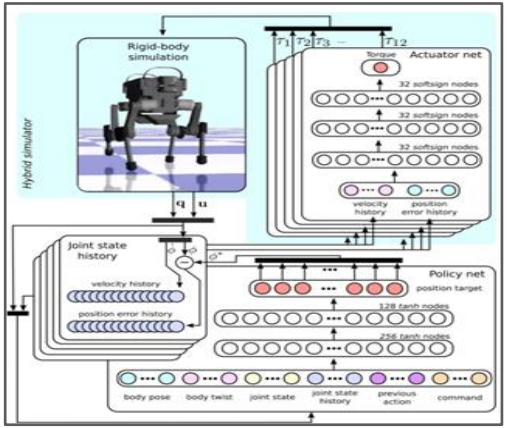
Previous Methods of Generating Gait Reward Functions

1. Use trajectory reference to specify gait rewards[1]



- Unadaptability in varying environment
- Challenging to derive reference trajectories

2. Reference-free approaches for specifying reward functions[2]



- Difficult to achieve specific gait characteristics
- Extending to other behaviors is hard

[1] H. Zhang, S. Starke, T. Komura, and J. Saito, "Mode-adaptive neural networks for quadruped motion control," ACM Transactions on Graphics, vol. 37, no. 4, pp. 1–11, 2018, ISSN: 15577368. DOI: 10.1145/3197517.3201366

[2] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," Science Robotics, vol. 4, no. 26, 2019.

Main Idea and Contribution

The paper introduces a periodic reward composition framework to enable the bipedal robot Cassie to learn and perform a wide range of gaits (e.g., walking, hopping, running) through sim-to-real reinforcement learning, without relying on predefined trajectories.

Contributions:

1. This work presents **a principled framework for designing reward functions** that can naturally capture all of the periodic bipedal locomotion gaits.
2. This work demonstrated this framework for **sim-to-real RL of all common bipedal gaits**, including walking, running, galloping, skipping, and hopping, **without using a motion capture dataset or reference trajectories**.

Key Formulas

Periodic Reward Composition

$$R(s,\phi) = \beta + \sum_i R_i(s,\phi)$$

$$R_i(s,\phi) = c_i \cdot I_i(\phi) \cdot q_i(s)$$

- c_i is a phase-specific coefficient controlling the weight of the reward for each phase.
- $I_i(\phi)$ is a phase indicator function that activates the reward depending on the current position in the gait cycle.
- $q_i(s)$ represents physical measurements (such as foot forces or velocities) relevant to the specific phase.

Phase Indicator Function

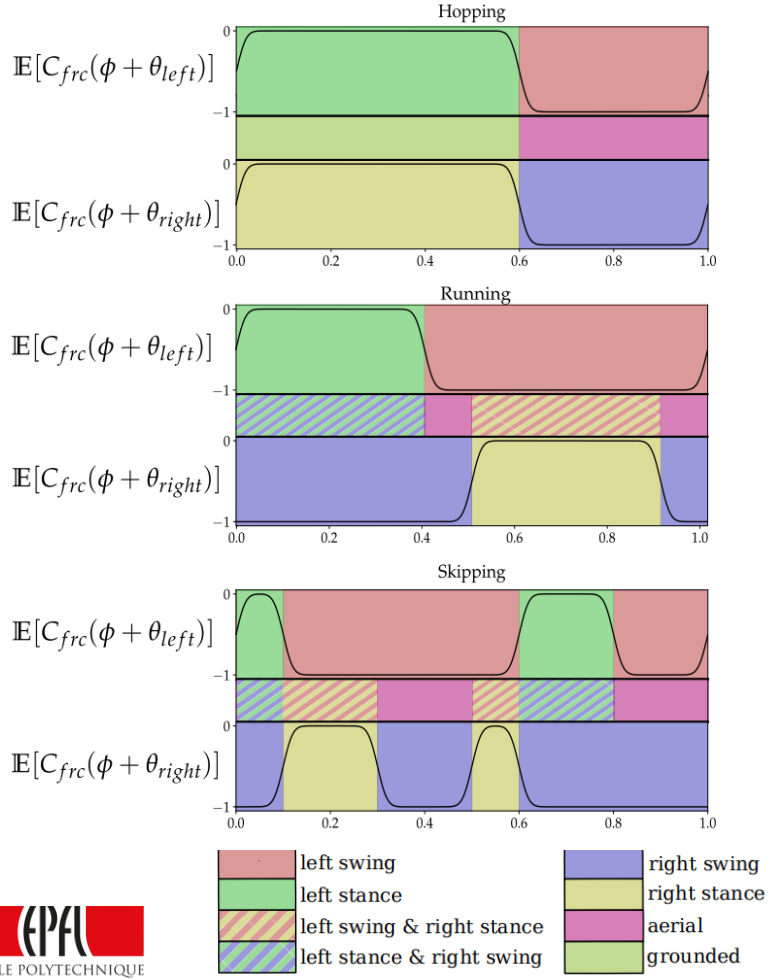
$$P(I_i(\phi) = x) = \begin{cases} P(A_i < \phi < B_i) & \text{if } x = 1 \\ 1 - P(A_i < \phi < B_i) & \text{if } x = 0 \end{cases}$$

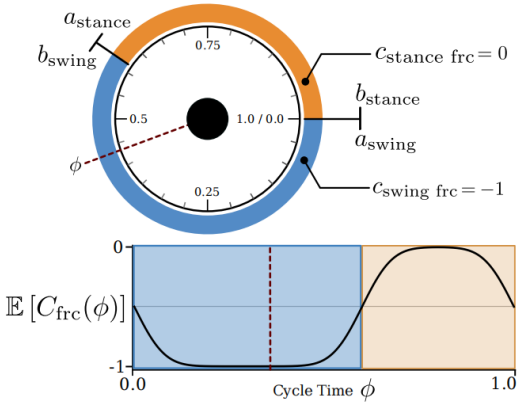
Here, A_i and B_i mark the start and end of the specific phase.

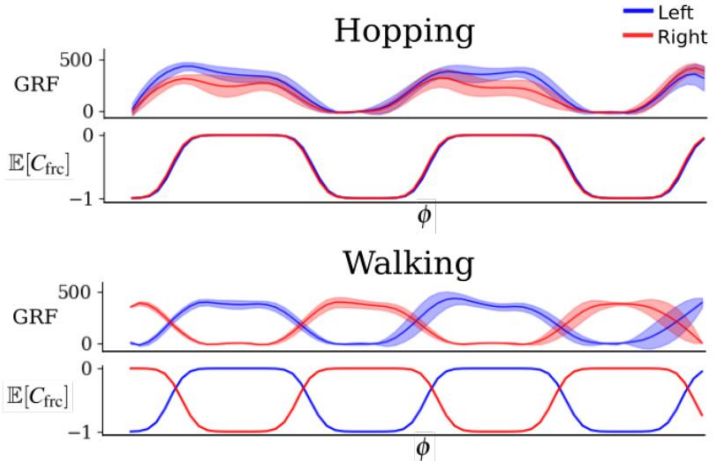
Bipedal Gait Reward Formula for Left and Right Feet

$$\begin{aligned} \mathbb{E}[R_{\text{bipedal}}(s,\hat{\phi})] = \quad & \mathbb{E}[C_{\text{frc}}(\phi + \theta_{\text{left}})] \cdot q_{\text{left frc}}(s) \\ & + \mathbb{E}[C_{\text{frc}}(\phi + \theta_{\text{right}})] \cdot q_{\text{right frc}}(s) \\ & + \mathbb{E}[C_{\text{spd}}(\phi + \theta_{\text{left}})] \cdot q_{\text{left spd}}(s) \\ & + \mathbb{E}[C_{\text{spd}}(\phi + \theta_{\text{right}})] \cdot q_{\text{right spd}}(s) \end{aligned}$$

- C_{frc} and C_{spd} represent phase reward functions, assessing foot forces and velocities respectively.
- θ_{left} and θ_{right} are the phase offsets for the left and right feet, ensuring alternating steps.
- $q_{\text{left frc}}(s)$ and $q_{\text{right frc}}(s)$ measure the foot forces for each foot, while $q_{\text{left spd}}(s)$ and $q_{\text{right spd}}(s)$ measure the velocities.


Hopping
$\mathbb{E}[C_{frc}(\phi+\theta_{left})]$
$\mathbb{E}[C_{frc}(\phi+\theta_{right})]$
Running
$\mathbb{E}[C_{frc}(\phi+\theta_{left})]$
$\mathbb{E}[C_{frc}(\phi+\theta_{right})]$
Skipping
$\mathbb{E}[C_{frc}(\phi+\theta_{left})]$
$\mathbb{E}[C_{frc}(\phi+\theta_{right})]$
0.0 0.2 0.4 0.6 0.8 1.0
left swing
left stance
left swing & right stance
left stance & right swing
right swing
right stance
aerial
grounded


a_{stance}
b_{swing}
0.75
$c_{\text{stance frc}} = 0$
b_{stance}
0.5
1.0 / 0.0
a_{swing}
ϕ
0.25
$c_{\text{swing frc}} = -1$
$\mathbb{E}[C_{\text{frc}}(\phi)]$
0.0
Cycle Time ϕ
1.0


Hopping
Left
Right
GRF
500
0
$\mathbb{E}[C_{\text{frc}}]$
ϕ
Walking
GRF
$\mathbb{E}[C_{\text{frc}}]$
ϕ

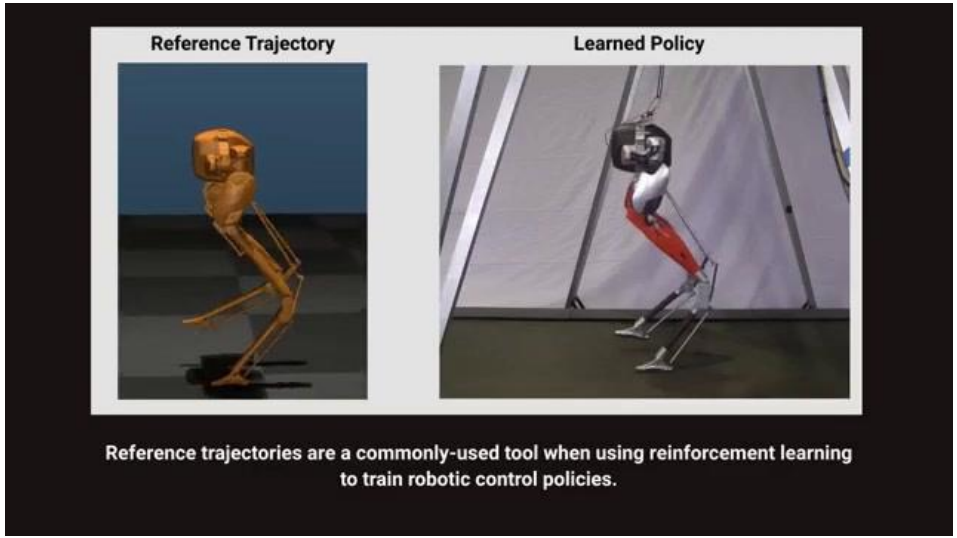
Result

Single-Gait Policy Results: The framework successfully trained policies Single-Gait by holding specific parameters constant.

Multi-Gait Policy Results: Training policies for smooth transitions between gaits (like hopping to running) proved challenging. Adjusting both cycle offsets and phase ratios sometimes led to undesired behaviors, such as asymmetrical gaits or mixed behaviors.

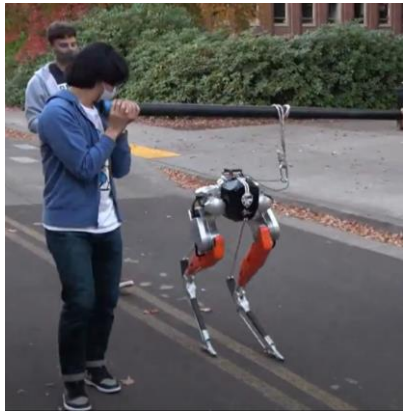
Solution with Transition Penalties:"transition penalties" are specific rewards or costs added to distinguish different behaviors during training, ensuring smoother transitions.

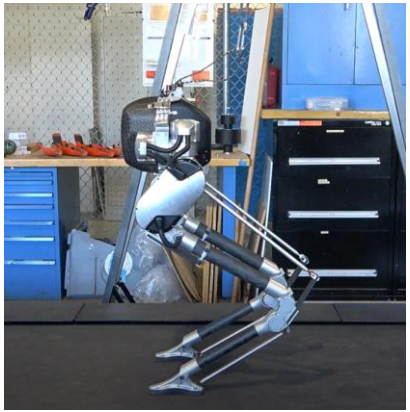
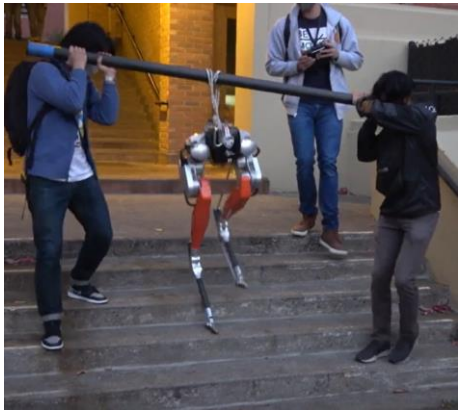
Experiment Background and Objectives



Since the robot lacks external sensors, it relies entirely on its internal states for adjusting its gait, which is effectively ‘blind’

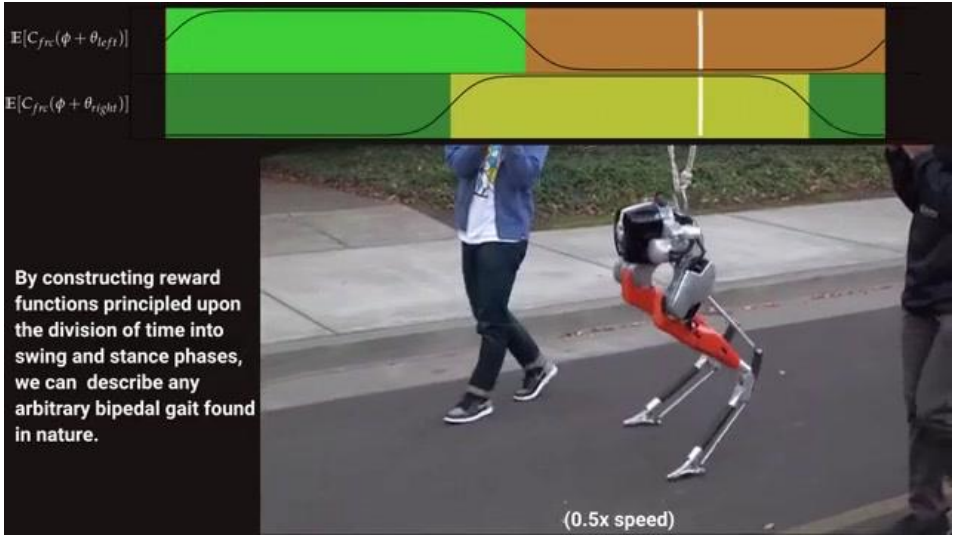
Experiment



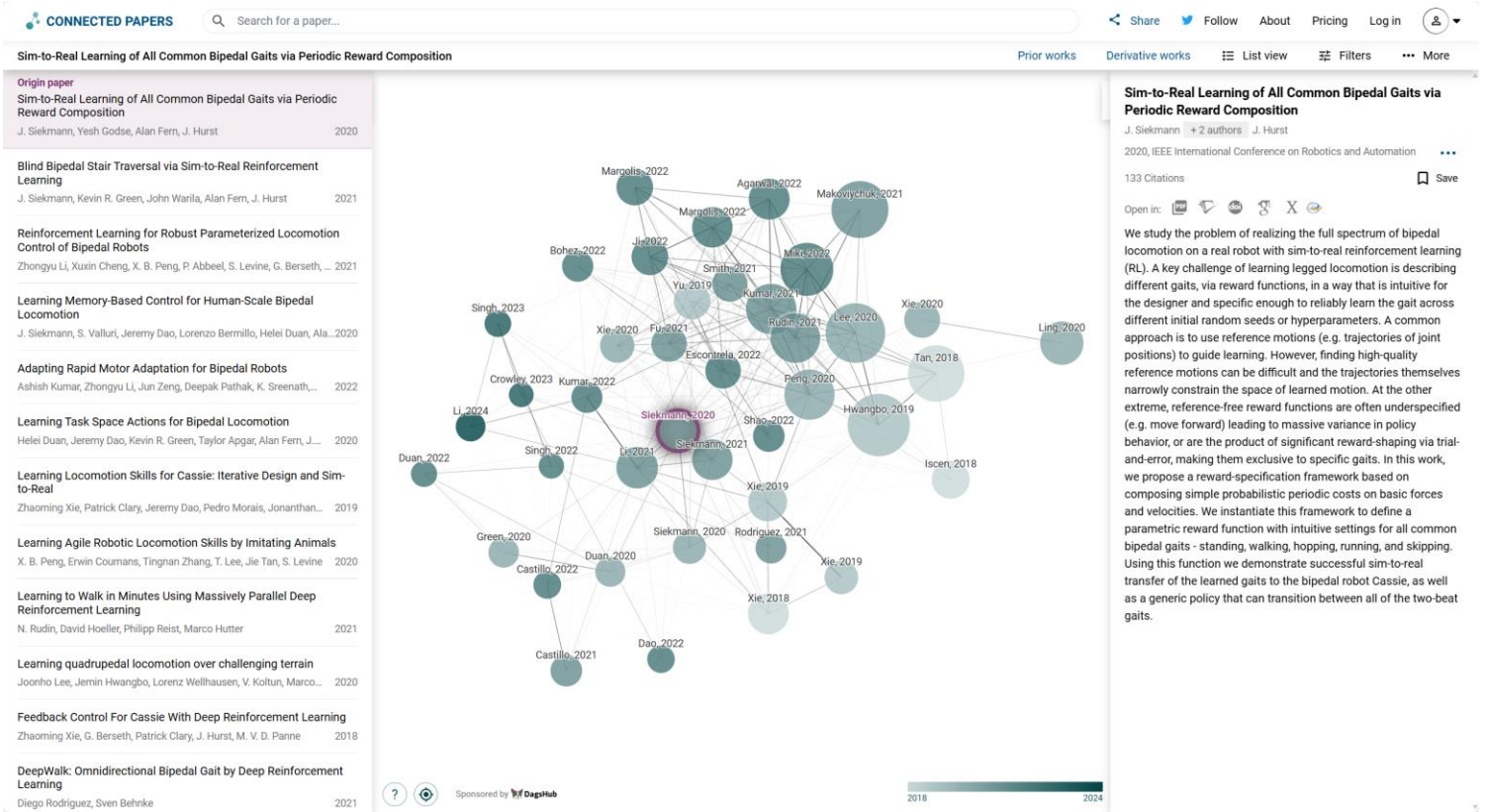
Tests involved the bipedal robot Cassie in environments like sidewalks, curbs, slopes, and stairs.

Experiment



Smooth transitions between gaits, such as moving from walking to running on a road

Connection


CONNECTED PAPERS
Search for a paper...
Share
Follow
About
Pricing
Log in
Sim-to-Real Learning of All Common Bipedal Gaits via Periodic Reward Composition
Prior works
Derivative works
List view
Filters
More
Origin paper
Sim-to-Real Learning of All Common Bipedal Gaits via Periodic Reward Composition
J. Siekmann, Yesh Godse, Alan Fern, J. Hurst 2020
Blind Bipedal Stair Traversal via Sim-to-Real Reinforcement Learning
J. Siekmann, Kevin R. Green, John Warila, Alan Fern, J. Hurst 2021
Reinforcement Learning for Robust Parameterized Locomotion Control of Bipedal Robots
Zhongyu Li, Xuxin Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, ... 2021
Learning Memory-Based Control for Human-Scale Bipedal Locomotion
J. Siekmann, S. Valluri, Jeremy Dao, Lorenzo Bermillo, Helei Duan, Ala... 2020
Adapting Rapid Motor Adaptation for Bipedal Robots
Ashish Kumar, Zhongyu Li, Jun Zeng, Deepak Pathak, K. Sreenath,... 2022
Learning Task Space Actions for Bipedal Locomotion
Helei Duan, Jeremy Dao, Kevin R. Green, Taylor Apgar, Alan Fern, J.... 2020
Learning Locomotion Skills for Cassie: Iterative Design and Sim-to-Real
Zhaoming Xie, Patrick Clary, Jeremy Dao, Pedro Morais, Jonanthan... 2019
Learning Agile Robotic Locomotion Skills by Imitating Animals
X. B. Peng, Erwin Coumans, Tingnan Zhang, T. Lee, Jie Tan, S. Levine 2020
Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning
N. Rudin, David Hoeller, Philipp Reist, Marco Hutter 2021
Learning quadrupedal locomotion over challenging terrain
Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, V. Koltun, Marco... 2020
Feedback Control For Cassie With Deep Reinforcement Learning
Zhaoming Xie, G. Berseth, Patrick Clary, J. Hurst, M. V. D. Panne 2018
DeepWalk: Omnidirectional Bipedal Gait by Deep Reinforcement Learning
Diego Rodriguez, Sven Behnke 2021
Margolis, 2022
Agarwal, 2022
Makoviychuk, 2021
Margolis, 2022
Bohez, 2022
Ji, 2022
Mikl, 2022
Smith, 2021
Yu, 2019
Kumar, 2021
Singh, 2023
Xie, 2020
Xie, 2020
Fu, 2021
Rudin, 2021
Lee, 2020
Ling, 2020
Escontrela, 2022
Tan, 2018
Crowley, 2023
Kumar, 2022
Peng, 2020
Li, 2024
Siekmann, 2020
Shao, 2022
Hwangbo, 2019
Duan, 2022
Singh, 2022
Li, 2021
Siekmann, 2021
Iscen, 2018
Xie, 2019
Green, 2020
Siekmann, 2020
Rodriguez, 2021
Duan, 2020
Xie, 2019
Castillo, 2022
Xie, 2018
Castillo, 2021
Dao, 2022
Sponsored by DagsHub
2018
2024
Sim-to-Real Learning of All Common Bipedal Gaits via Periodic Reward Composition
J. Siekmann +2 authors J. Hurst
2020, IEEE International Conference on Robotics and Automation
133 Citations
Save
Open in:
We study the problem of realizing the full spectrum of bipedal locomotion on a real robot with sim-to-real reinforcement learning (RL). A key challenge of learning legged locomotion is describing different gaits, via reward functions, in a way that is intuitive for the designer and specific enough to reliably learn the gait across different initial random seeds or hyperparameters. A common approach is to use reference motions (e.g. trajectories of joint positions) to guide learning. However, finding high-quality reference motions can be difficult and the trajectories themselves narrowly constrain the space of learned motion. At the other extreme, reference-free reward functions are often underspecified (e.g. move forward) leading to massive variance in policy behavior, or are the product of significant reward-shaping via trial-and-error, making them exclusive to specific gaits. In this work, we propose a reward-specification framework based on composing simple probabilistic periodic costs on basic forces and velocities. We instantiate this framework to define a parametric reward function with intuitive settings for all common bipedal gaits - standing, walking, hopping, running, and skipping. Using this function we demonstrate successful sim-to-real transfer of the learned gaits to the bipedal robot Cassie, as well as a generic policy that can transition between all of the two-beat gaits.

Connection

biomimetics

MDPI

Article

Adaptive Gait Acquisition through Learning Dynamic Stimulus Instinct of Bipedal Robot

Yuanxi Zhang [1], Xuechao Chen [1,2], Fei Meng [1,2,*], Zhangguo Yu [1,2], Yidong Du [1], Zishun Zhou [1] and Junyao Gao [1,2]

[1] School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China; zhangyuanxi@bit.edu.cn (Y.Z.); chenxuechao@bit.edu.cn (X.C.); yuzg@bit.edu.cn (Z.Y.); duyidong@bit.edu.cn (Y.D.); zishun.zhou@outlook.com (Z.Z.); gaojunyao@bit.edu.cn (J.G.)
[2] Key Laboratory of Biomimetic Robots and Systems, Ministry of Education, Beijing 100081, China
* Correspondence: mfly0208@bit.edu.cn

Abstract: Standard alternating leg motions serve as the foundation for simple bipedal gaits, and the effectiveness of the fixed stimulus signal has been proved in recent studies. However, in order to address perturbations and imbalances, robots require more dynamic gaits. In this paper, we introduce dynamic stimulus signals together with a bipedal locomotion policy into reinforcement learning (RL). Through the learned stimulus frequency policy, we induce the bipedal robot to obtain both three-dimensional (3D) locomotion and an adaptive gait under disturbance without relying on an explicit and model-based gait in both the training stage and deployment. In addition, a set of specialized reward functions focusing on reliable frequency reflections is used in our framework to ensure correspondence between locomotion features and the dynamic stimulus. Moreover, we demonstrate efficient sim-to-real transfer, making a bipedal robot called BITeno achieve robust locomotion and disturbance resistance, even in extreme situations of foot sliding in the real world. In detail, under a sudden change in torso velocity of −1.2 m/s in 0.65 s, the recovery time is within 1.5–2.0 s.

Keywords: reinforcement learning; bipedal robot; adaptive locomotion; period dynamic gait

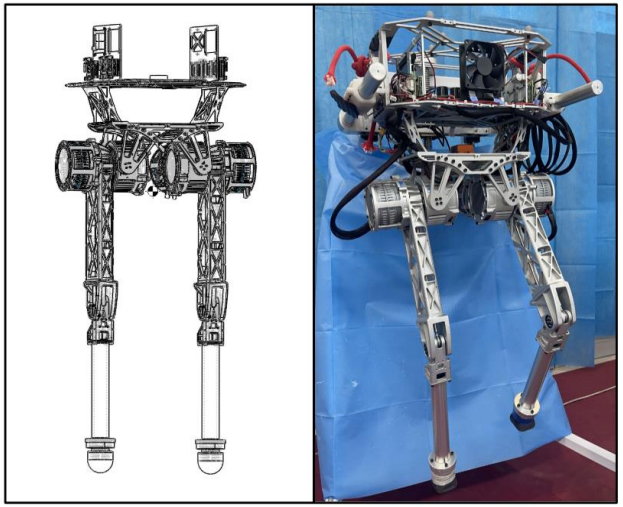


Figure 2. The design of BITeno platform. (**Left**) The mechanical design in simulation, were the feature of each link was assigned according to real materials. (**Right**) The physical robot with an electrical system onboard.

Pros & Cons

Pros

Versatile Gait Learning
Eliminates Dependency on Reference Trajectories
Adaptability to Environmental Disturbances

Cons

Limited Adaptability to High-Level Disturbances

Possible exam questions

·How can reward functions be designed to achieve different gait behaviors without reference trajectories?
·Explain the role of periodic reward composition in learning bipedal gaits for a robot.

Answer 1: Periodic reward composition allows for the specification of distinct gait characteristics by defining phases (swing and stance) with specific penalties or rewards. This framework guides the reinforcement learning process, enabling the robot to learn stable bipedal gaits like walking, running, hopping, and skipping by rewarding appropriate foot forces and velocities during each phase.

Answer 2: Without reference trajectories, reward functions can be designed using periodic reward composition. By assigning rewards or penalties to foot speed and force during different gait phases (such as penalizing foot force during the swing phase and penalizing foot speed during the stance phase), the framework encourages the robot to learn and achieve distinct gait characteristics without relying on specific motion trajectories.

Outdoor Experiments of Bipedal Robot Gait Learning

Aim to demonstrate the robot's stability and adaptability in real-world outdoor settings