Iterative Gait Design for Cassie using DeepRL

Zhaoming Xie*, Patrick Clary†, Jeremy Dao†, Pedro Morais†, Jonathan Hurst† and Michiel van de Panne*
*Department of Computer Science
University of British Columbia, Vancouver, BC, Canda
Email: {zxie47, van}@cs.ubc.ca
†Dynamic Robotics Laboratory
Oregon State University, Corvallis, Oregon, USA
Email: {claryp,daoje,autranep,jonathan.hurst}@oregonstate.edu

***Abstract*—Deep reinforcement learning (DRL) is a promising approach for developing legged locomotion skills. However, the iterative design process that is inevitable in practice is poorly supported In this paper, we propose a practical method that allows the reward function to be fully redefined on each successive design iteration while limiting the deviation from the previous iteration. We characterize policies via sets of Deterministic Action Stochastic State (DASS) tuples, which represent the deterministic policy state-action pairs as sampled from the states visited by the trained stochastic policy. New policies are trained using a policy gradient algorithm which then mixes RL-based policy gradients with gradient updates defined by the DASS tuples. The tuples also allow for robust policy distillation to new network architectures. We demonstrate the effectiveness of this iterative-design approach on the bipedal robot Cassie, achieving stable walking with different gait styles at various speeds.**

I. Introduction

In this paper, we propose a DRL design process that reflects and supports the iterative nature of control policy design. At its heart is a data collection technique that allows us to recover a trained policy from a relatively small number of samples. With this technique, we can quickly compress and combine locomotion policies with supervised learning. By using policy gradient updates that combine the supervised learning samples and conventional DRL policy-gradient samples, we allow for the iterative design of improved policies using new reward functions that encourage desired behaviors. We validate our approach in simulation and on a physical Cassie robot, demonstrating stable walking policies with different styles at various speeds. Frames from a learned forward-walking gait for Cassie are shown in Fig. 1. Video of Cassie walking with different policies can be seen at https://youtu.be/0cD8Y-1BkOU.

We believe this is the first time that neural network policies for variable-speed locomotion have been successfully deployed on a human-scale $3D$ bipedal robot such as Cassie. The policies trained in simulation are directly transferred to the physical robot without the use of the dynamics randomization methods. The gaits are comparable or faster in speed than other gaits reported in the literature for Cassie, e.g., [1].

II. Design Process

We develop our policies iteratively. Fig. 2 shows our design process. At each design iteration, designers come up with new reward function for desirable behaviors. DASS samples from previous iteration are used to guide the policy optimization.

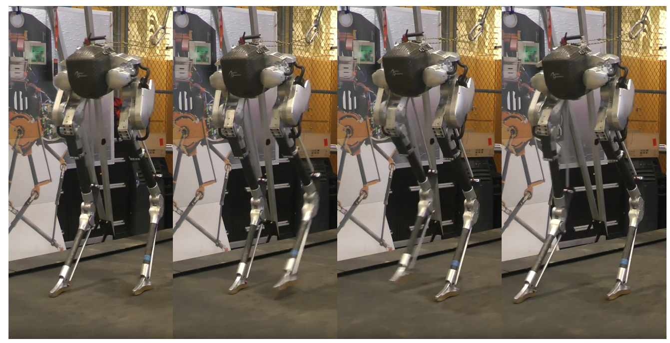

Fig. 1: Cassie walking on a treadmill with a neural network policy.

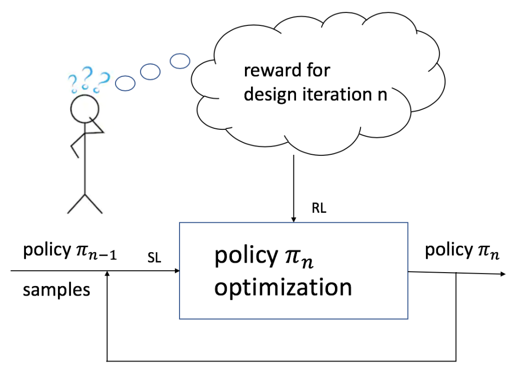

Fig. 2: Iterative RL-based design process. Policy π_n is optimized using policy gradients from both supervised learning (SL) and reinforcement learning (RL).

References

[1] Yukai Gong, Ross Hartley, Xingye Da, Ayonga Hereid, Omar Harib, Jiunn-Kai Huang, and Jessy W. Grizzle. Feedback control of a cassie bipedal robot: Walking, standing, and riding a segway. *CoRR*, abs/1809.07279, 2018. URL http://arxiv.org/abs/1809.07279.