A Review of Bipedal Locomotion using Reinforcement Learning

Presenter: Guojing Huang

2024/3/19

Motivation and Main Problem

The problem of 3D bipedal walking is important for general-purpose robot autonomy and has potential applications and societal impact. The main problem is to design feedback control policies for stable and robust 3D bipedal locomotion.

- Significance towards general-purpose robot autonomy :
  - Solving the problem of 3D bipedal walking enables the development of autonomous robots capable of walking in complex environments.
  - This has potential applications in areas such as search and rescue, disaster response, and personal assistance robotics.

- Technical Challenges:
  - 3D bipedal walking is a challenging problem due to the multi-phase and hybrid nature of legged locomotion, underactuation, unilateral ground contacts, nonlinear dynamics, and high degrees of freedom.
  - Existing approaches, both model-based and model-free, have limitations in capturing the complex dynamics of a 3D robot in the real world, resulting in non-robust controllers that require additional heuristic compensations and tuning processes.

SUSTech

AncoraSIR.com

Motivation and Main Problem

The problem of 3D bipedal walking is important for general-purpose robot autonomy and has potential applications and societal impact. The main problem is to design feedback control policies for stable and robust 3D bipedal locomotion.

- Role of AI and Machine Learning:
  - The proposed solution utilizes a model-free reinforcement learning (RL) framework to design feedback control policies for 3D bipedal walking.
  - The RL framework incorporates physical insights gained from the hybrid nature of the walking dynamics and the hybrid zero dynamics approach, improving the data efficiency of the RL method and reducing the need for manual tuning of feedback regulations.
- Reasons for Prior Approaches Not Solving the Problem:
  - Prior approaches, both model-based and model-free, have limitations in capturing the complex dynamics of a 3D robot in the real world, resulting in non-robust controllers that require additional heuristic compensations and tuning processes.

AncoraSIR.com

Problem Setting

- The problem is the design of feedback control policies for stable and robust 3D bipedal locomotion using reinforcement learning (RL)
- The goal is to develop a unified policy that can handle trajectory planning and feedback control for 3D walking
- The problem is challenging due to the multi-phase and hybrid nature of legged locomotion, underactuation, unilateral ground contacts, nonlinear dynamics, and high degrees of freedom
- Existing approaches, both model-based and model-free, have limitations in capturing the complex dynamics of a 3D robot in the real world, resulting in non-robust controllers that require additional heuristic compensations and tuning processes .

AncoraSIR.com

Context / Related Work / Limitations of Prior Work

Model-based Methods

1. Limitation of mathematical methods

- Difficulty in capture the complex dynamics of a 3D robot in the real world.
- Time-consuming & requires experiences: the non-robust controllers require additional heuristic compensations and tuning processes.

2. Reduced order models

- Rely on strong assumptions: quasi-static and unrealistic walking behaviors.
- Computationally expensive.
- Sensitive to model parameters and environmental changes.

AncoraSIR.com

Context / Related Work / Limitations of Prior Work

Model-free methods (Reinforcement Learning)

1. Simple 2D robot model (End-to-end training)

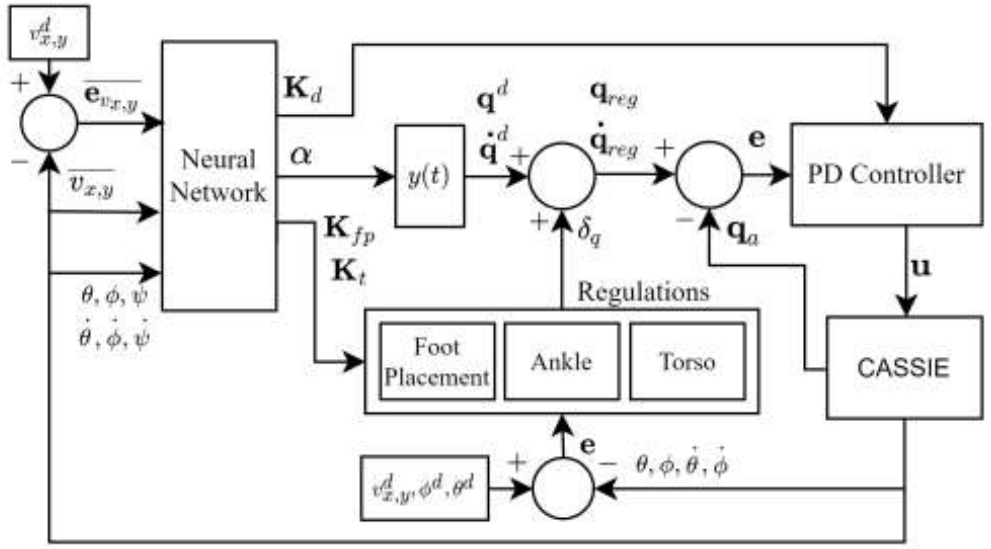
- Sampling inefficient
- Over-parameterized
- Non-smooth control signals and unnatural motions that are not applicable to real robots

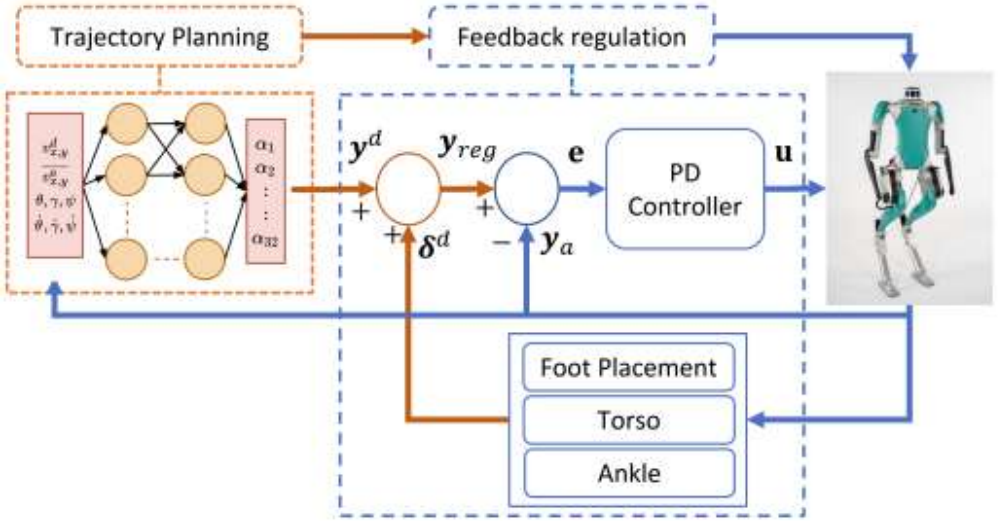
2. More complex 3D bipedal walking (RL methods as part of the feedback control)

AncoraSIR.com

Method

Overall RL framework





·Optimization objective: the walking gait

·mapping tool: a NN function

·parameters: 1) a set of coefficients of the Bezier polynomials 2) a set of gains corresponding to the derivative 3) gain of the joints PD controller 4) the gains for the foot placement 5) the gain for the torso regulations

·Breakthrough: MODEL-FREE

the reference trajectories are learned from scratch > rely on given working policy providing the joints reference trajectories

·What is it: a non-conventional RL framework

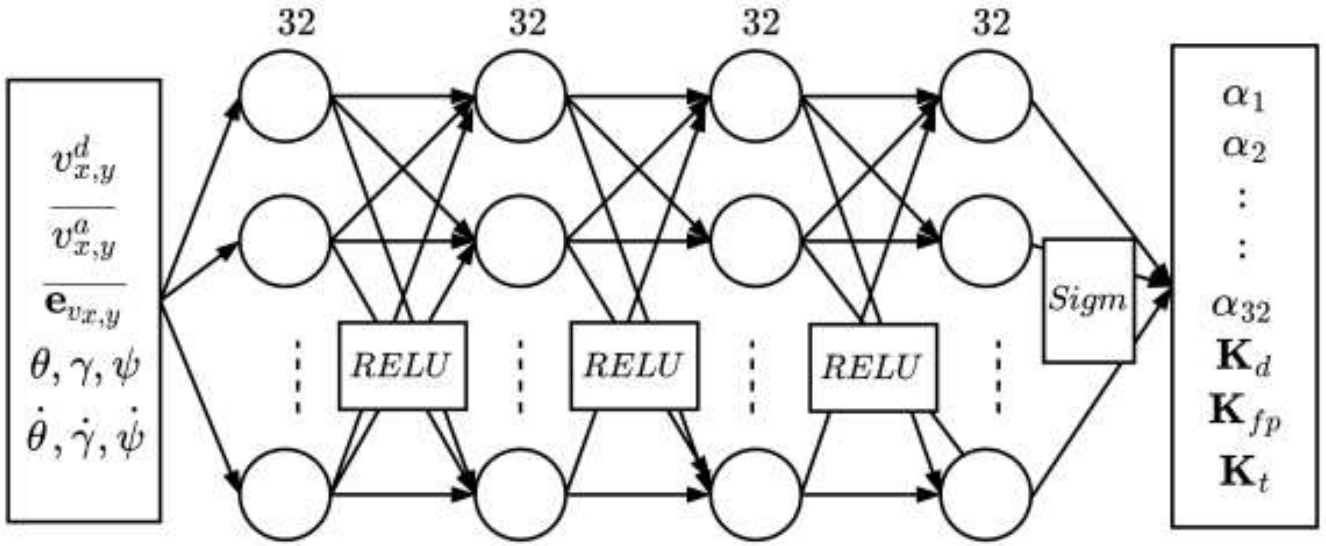
AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Method

The neural network structure



$$q_i^{min} < \alpha_i^{min} < q_i^d < \alpha_i^{max} < q_i^{max}.$$

·Optimization objective: the walking gait
·What is it: non-conventional RL framework
·Inputs layer: velocity, error, angles, angular velocities
·4 hidden layers, 4x32 neurons
·RELU activation functions(线性整流函数): solving nonlinear problem, improve computational efficiency
·Sigmoid function: binary classification
·Output layers:
1) the coefficients of the Bezier polynomials α,
2) the set of gains of the PD controller, foot placement and torso compensations, denoted by Kd, Kf p and Kt

·Boundary: decreases the complexity

AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Theory

Hybrid Zero Dynamics(HZD)

The HZD approach is based on the concept of virtual constraints, which are mathematical relationships that define desired behaviors or constraints for the robot's motion.

$$\mathbf{y}_2 := \mathbf{y}_2^a(\mathbf{q}) - \mathbf{y}_2^d(\tau(t), \alpha),$$

$$\mathbf{y}_2^d(\tau(t), \alpha) := \sum^{5} \alpha[k] \frac{M!}{k!(M-k)!} \tau(t)^k (1-\tau(t))^{M-k}.$$

·Virtual Constraints

$$\tau(t) = \frac{t - t^-}{t_{step}},$$

·a vector of desired outputs defined in terms of 5th order Bezier polynomials parameterized by the coefficients α

·the scaled relative time with respect to step time interval

$$\delta^{sw}_{hpitch}[k] = K_{p_x}(v_x[k] - v_x^d) + K_{d_x}(v_x[k] - v_x[k-1]),$$
$$\delta^{sw}_{hroll}[k] = K_{p_y}(v_y[k] - v_y^d) + K_{d_y}(v_y[k] - v_y[k-1]).$$

·Foot placement controller

$$u^{st}_{hroll} = K_{p_{troll}}(\phi - \phi^d) + K_{d_{troll}}(\dot{\phi} - \dot{\phi}^d),$$
$$u^{st}_{hpitch} = K_{p_{tpitch}}(\theta - \theta^d) + K_{d_{tpitch}}(\dot{\theta} - \dot{\theta}^d),$$

·Torso regulation

$$\gamma^{sw} = \theta - 13\,\text{deg} - 50\,\text{deg},$$

·Ankle regulation

AncoraSIR.com

Experimental Setup

Experimental Setup

- A customized environment for Cassie was built using Mujoco
- The model information of Cassie robot provided by Agility Robotics was used
- The number of trainable parameters for the NN is 5069
- The training time is about 10 hours using a single 12-core CPU machine

- The performance of the control policy obtained from the training are presented in terms of
  (i) speed tracking,
  (ii) disturbance rejection,
  (iii) the convergence of stable periodic limit cycles.

AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Experimental Results

Speed Tracking

The performance of tracking a fixed desired speed of 0.5 m=s in the forward direction is shown in Fig. 5.

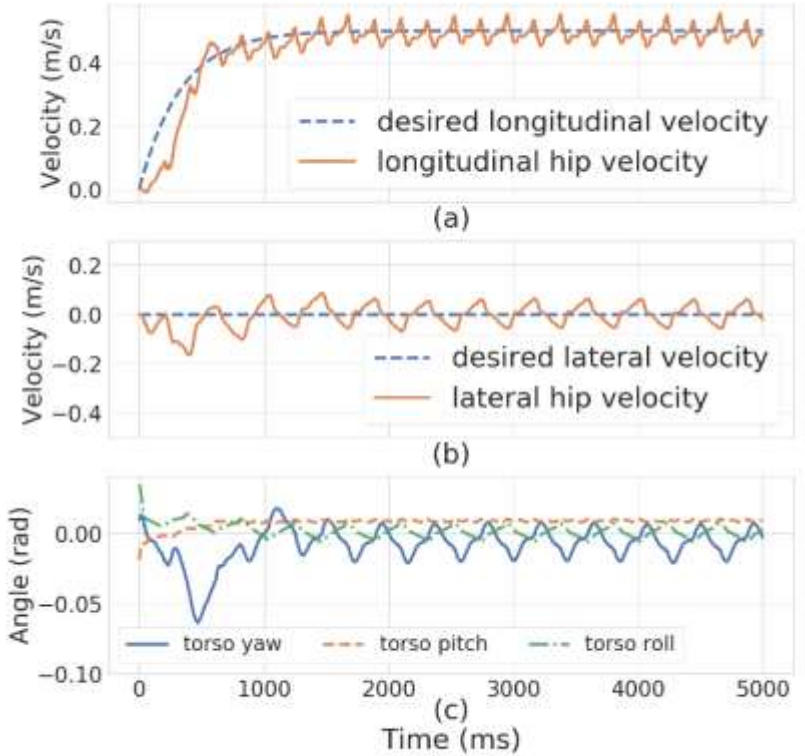


Fig. 5: Performance of the learned policy while tracking a fixed desired longitudinal walking speed.

AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Experimental Results

Speed Tracking

The performance of continuously tracking various desired speeds is shown in Fig. 6, in an interval from -0.5 to 1.0 m=s longitudinally (vx), and an interval from -0.3 to 0.3 m=s in the lateral direction (vy).

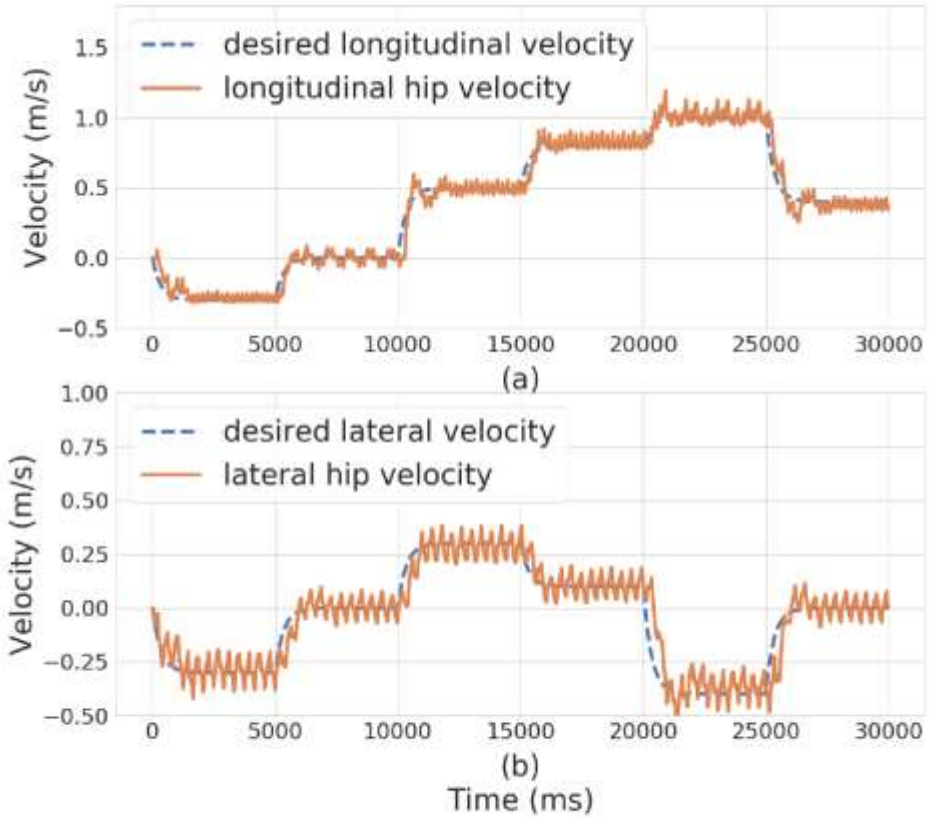


Fig. 6: Performance of the learned policy while tracking varying desired longitudinal and lateral walking speeds.

AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Experimental Results

Disturbance Rejection

To evaluate the robustness of our controller, they applied an adversarial force directly at the robot's pelvis in both the forward and the backward directions.

In the results shown in Fig. 7 and Fig. 8, they adopt the adversarial force with the same magnitude of 25 N in both directions. Itis applied 2 seconds after starting the test and lasts for 0.1 seconds. Throughout our tests, the robot can handle up to 40 N in the forward direction and 45 N in the backward direction without falling, but the speed tracking may take a long time to recover with an external force of high magnitude.

AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Experimental Results

Disturbance Rejection

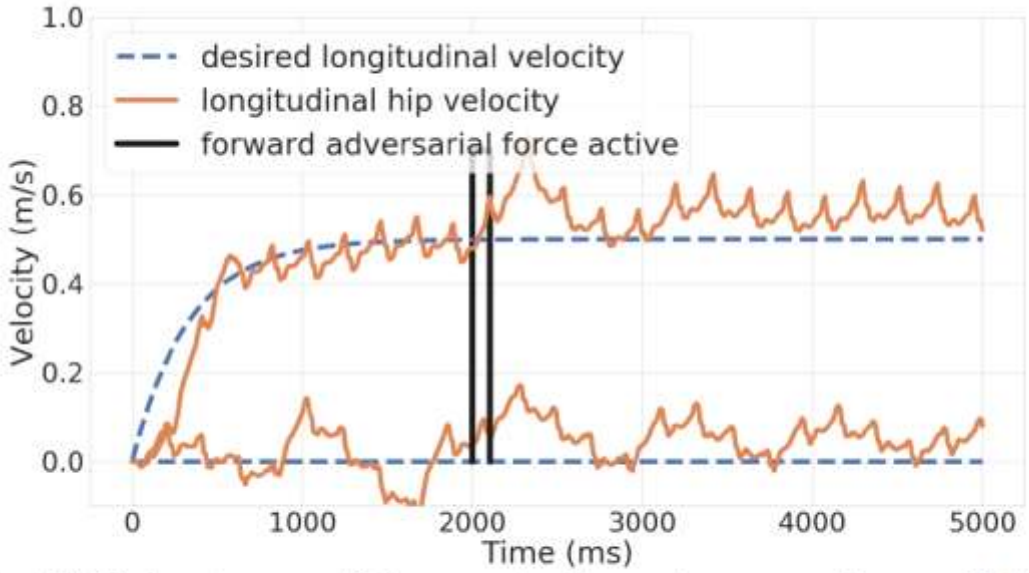


Fig. 7: Robustness of the controller when an adversarial force is applied in the forward direction.

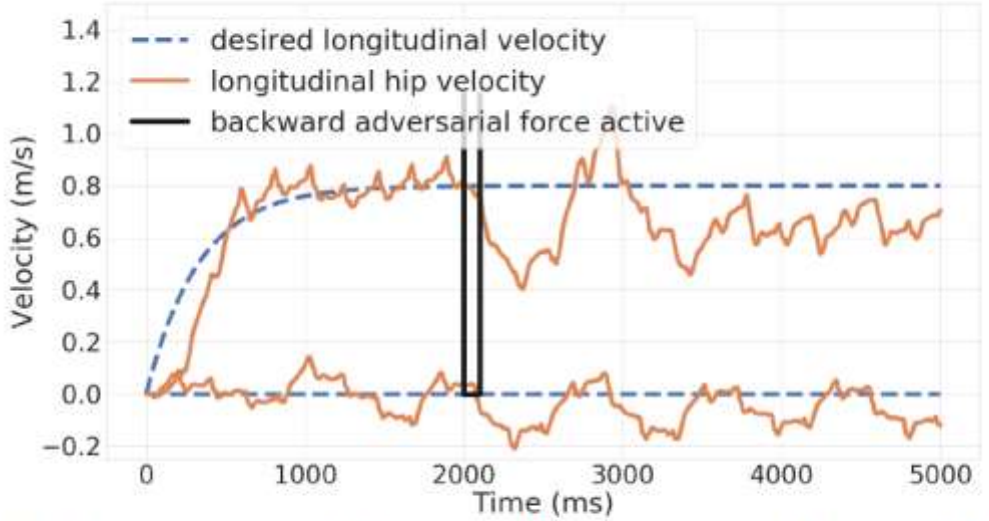


Fig. 8: Robustness of the controller when an adversarial force is applied in the backward direction.

AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Experimental Results

Periodic Stability of the Walking Gaits

Fig. 9 shows that the convergence of several representative robot actuated joints to periodic limit cycles during a fixed speed walking. Moreover, the orbit described by the left and right joints demonstrates the symmetry of walking gaits.

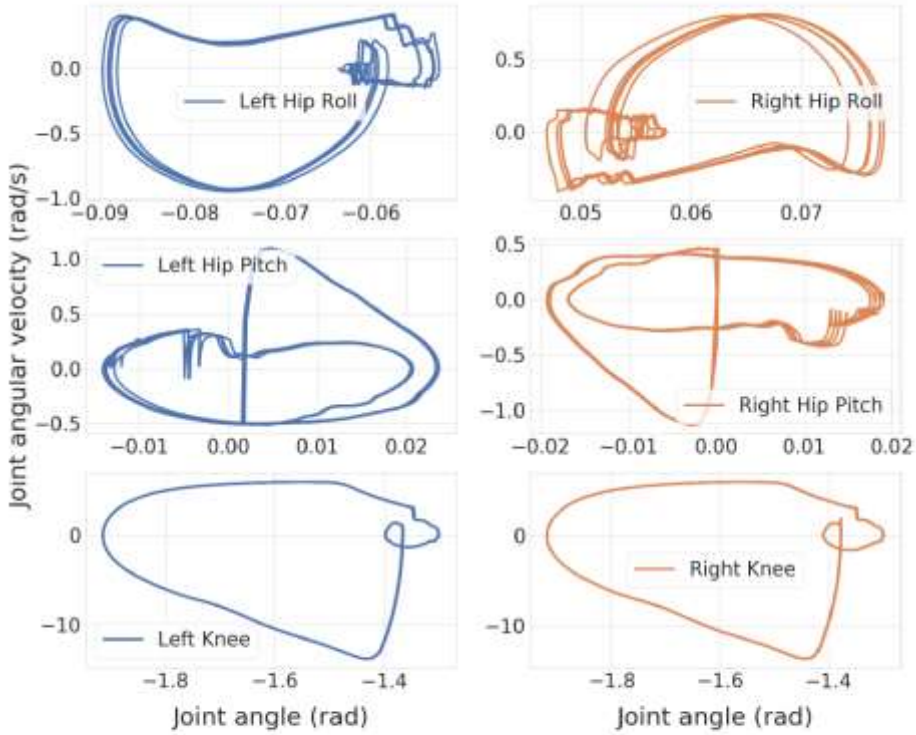


Fig. 9: Walking limit cycle of the learned policy with the desired longitudinal velocity of 0.5 m/s.

AncoraSIR.com

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Hybrid Zero Dynamics Inspired Feedback Control Policy Design for 3D Bipedal Locomotion using Reinforcement Learning," 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 8746-8752.

SUSTech

Discussion of Results

Insights gained form the results

- decoupled structure → effective speed tracking in both longitudinal and lateral directions → overall stability of the walking gait
- disturbance rejection tests → confirm robustness
- capability to achieve periodic stability in walking gaits → potential for practical implementation in real-world scenarios

Supporting evidences

- Evidence of the policy's effectiveness in speed tracking, disturbance rejection, and achieving periodic stability.
- Data efficiency and reduced neural network complexity

Additional experiments can further support conclusions

- Comparative analysis with traditional control methods to evaluate the proposed framework's performance relative to established techniques.

AncoraSIR.com

SUSTech

Discussion of Results

Strengths and weaknesses

1. Strengths

- Reduced number of neural network parameters → Efficiency and simplicity
- handling various walking speeds and disturbances → Flexibility and adaptability
- novel integration of physical walking dynamics insights into the model-free RL framework

2. Weaknesses

- Reliance on simulation results, lacking real-world validation.
- No comparison with traditional control techniques.

AncoraSIR.com

SUSTech

Limitations / Open Issues

Noise

The simulating duration contains no noise, which is impossible in real time condition. This paper failed to examine the performance under conditions filled noise.

Limited Disturbance Rejection

This simulating result performs a good resistance when a force of 25N applied on its backward direction, but only this performance is presented in details. What if the disturbance is more complex?

AncoraSIR.com

Limitations / Open Issues

Model Free

“Model Free” can be a “double-edged sword”, although this paper does not implicate the short come of the model, we are to consider the limitation on low efficiency, over-parameterlization, and interpretability, corresponding to the universality of the model, determining where to be applied.

Adaption to Real World

This paper focus on simulation, generalizations to the real world need to be examined.

AncoraSIR.com

Future Work for Paper

Noise and Disturbance injection

Injecting noise to the sensors and actuators during the training process, to enhance the robustness of the learned policy.

Besides, apply complicated disturbance to examine and improve the generazation of the model, to better the simulation result of our own works.

Simplification

Simplify the abundant part of the model, to better the adaptability to our simulating, and in addition to improve the universality and interpretability.

AncoraSIR.com

Extended Readings

Robust Feedback Motion Policy Design Using Reinforcement Learning on a 3D Digit Bipedal Robot

This paper presents a framework for learning robust bipedal locomotion policies that can be transferred to real hardware with minimal tuning. By combining a sample effi cient learning structure with intuitive but powerful feedback regulations in a cascade structure, we decouple the learning problem into two stages that work at a different frequency to facilitate the implementation of the controller in the real hardware.

Reference: G. A. Castillo, B. Weng, W. Zhang and A. Hereid, "Robust Feedback Motion Policy Design Using Reinforcement Learning on a 3D Digit Bipedal Robot," 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021, pp. 5136-5143.

AncoraSIR.com

Extended Readings

Learning robust perceptive locomotion for quadrupedal robots in the wild

This essay presents a robust and general solution to integrating exteroceptive and proprioceptive perception for legged locomotion. They leverage an attention-based recurrent encoder that integrates proprioceptive and exteroceptive input. The encoder is trained end to end and learns to seamlessly combine the different perception modalities without resorting to heuristics. The result is a legged locomotion controller with high robustness and speed.

Reference: Takahiro Miki et al. ,Learning robust perceptive locomotion for quadrupedal robots in the wild.Sci. Robot.7,eabk2822(2022).

AncoraSIR.com

SUSTech

Summary

Problem the reading is discussing

This paper presents a novel model-free RL approach for the design of feedback controllers for 3D bipedal robots.

Importance and difficulties

The main contribution of this work does not focus on the control algorithm but on a novel RL framework with enhanced features over traditional RL methods.

Key limitation of prior work

Existing RL algorithms are often trained in an end-to-end manner or rely on prior knowledge of some reference joint trajectories.

AncoraSIR.com

Summary

Key insights of the proposed work

- Model-free Framework for 3D Bipedal Locomotion
- Efficiency since less trainable parameters
- Robustness evaluated by several disturbance rejection tests

What did authors demonstrate by these insights

The result is a data-efficient RL method with a reduced number of parameters in the NN that can learn stable and robust dynamic walking gaits from scratch, without any reference motion or expert guidance.

AncoraSIR.com

A Review of Bipedal Locomotion using Reinforcement Learning

Guojing Huang
Robotics Engineering
Chaoyang Song
12111820@mail.sustech.edu.cn

AncoraSIR.com

SUSTech