Title	Deep Reinforcement Learning for Tailorable Natural Quadruped Gait Generation
Author(s)	Sulpice, Lucas; Owaki, Dai; Hayashibe, Mitsuhiro
Citation	The 11th International Symposium on Adaptive Motion of Animals and Machines (AMAM2023). 2023, p. 87-88
Version Type	VoR
URL	https://doi.org/10.18910/92280
rights	
Note	

Deep Reinforcement Learning for Tailorable Natural Quadruped Gait Generation

Lucas Sulpice[1], Dai Owaki[1], Mitsuhiro Hayashibe[1]
[1]Neuro-Robotics Lab, Department of Robotics, Graduate School of Engineering, Tohoku University, Japan
sulpice@dc.tohoku.ac.jp — owaki@tohoku.ac.jp — hayashibe@tohoku.ac.jp

1 Introduction

Quadruped robots and models have become common and widely researched subjects in the past years [1]. These studies have been conducted to achieve locomotion of quadruped robots at various speed [2], using various gaits [2] and how they adapt their gait to the terrain [3], leading to a better understanding of quadruped locomotion in animals. In order to avoid using hand-crafted controllers to generate the studied gait types, the use of Deep Reinforcement Learning (DRL) to generate quadruped locomotion has been attracting attention. Although the locomotion is successfully generated, most of the trained models do not consistently achieve to learn a natural and well-known gait pattern. Recently, the idea of implementing conditions on the model to generate a specific gait has been introduced by imposing conditions on the actions of the model's joints [4]. Related to this idea of introducing new ways to ensure specific gaits in quadruped model trained by DRL, the purpose of this article is to:

- Generate a deterministic gait type by introducing conditions on the step order.
- Analyze the energetic properties of the learned gait types and show their similarities with quadruped animal gaits.

2 Methods

In order to simulate a realistic environment, we used the Mujoco physic engine [5], a classic and famous engine to simulate physical properties. For the study, we used our version of a simplistic quadruped Full-Cheetah model (Fig.1) having a rigid body and 12 joints. To make the quadruped model generate locomotion, we trained it using the Soft Actor Critic algorithm [6] as it achieved the best results for the model.

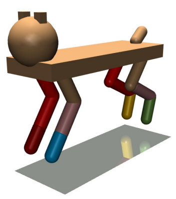

Figure 1: Quadruped Full-Cheetah model.

The model needs to reach a specific velocity, using a specific gait, while being energy optimized. We introduced both those parameters in the reward function (Eq.1).

$$R = -5\left|\frac{V(t)-V_t}{V_t}\right| - 0.1\sum_i A_i(t)^2 + StepReward(t) \quad (1)$$

The first term of the reward function is the relative difference between the current velocity $V(t)$ and the target velocity V_t and allows the model to reach the desired speed. The second term is the sum of the actions $A(i)$ of the model, the actions are the magnitude of the torques of each joints and represents the global effort for the model to move. Inspired by curriculum learning [7], this part of the reward function is inactive for the first third of the training to avoid the model converging to the sub-optimal solution of not moving at all. The third term is the StepReward, the calculation of this part of the reward depends on desired gait (B/F R/L stands for the Back/Front Right/Left leg; *nb_conctact* is the number of legs in contact with the ground):

- No Imposed Gaits (Full-Cheetah, FC): 0.
- Gallop: -1 if FR/FL on ground and BR/BL on ground.
- Trot: -1 if $nb\_contact! = \{0;2\}$ or [FR/BL on ground and FL/BR on ground].
- Walk: -1 if $nb\_contact! = 3$ or if the new step does not match the previous step (i.e wrong step order). The order of the moved legs is FL - BR - FR - BL.

These StepRewards were chosen to guide the model towards the desired gait. Animal-like gaits step order [8] are represented in Figure 2.

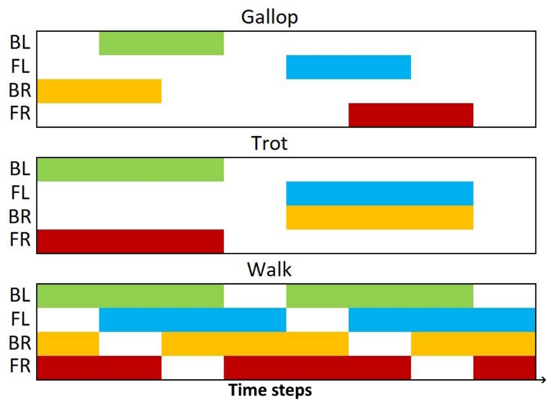

Figure 2: Step order for Gallop, Trot and Walk gaits.

In order to measure the performance of the model, a Performance index (Eq.2) is introduced.

$$Perf = \frac{Velocity}{Actions} = \frac{\sum_t V(t) * \Delta t}{\sum_t \sum_i A_i^2} \quad (2)$$

This indicator takes the velocity of the model and divides it by the sum of the squares of the actions for one full episode. This index is optimized by the reward function and represent the performance of the current gait at the current speed.

3 Results

The model was trained for four different gaits : no imposed gaits (FC), Gallop, Trot and Walk; and for six different target velocities: $1, 1.5, 2, 2.5, 3, 3.5 m.s^{-1}$. Each of the models was trained multiple times and the ones achieving the best StepReward and Performance were kept. However for the walk gait, velocities greater than $2 m.s^{-1}$ could not achieve high StepReward due to the non capability to reach these velocities while walking. Videos of the trained model are available: `youtu.be/aa23SG6Ho74`.

3.1 Generated gaits

In order to analyze the gaits generated by the trainings and the validity of those gaits, we visualized the contacts patterns between the feet and the ground for 100 steps in Figure 3. For each gait, we look at the one achieving the best performance (i.e Gallop 2.5 $m.s^{-1}$, Trot $2 m.s^{-1}$, Walk $1 m.s^{-1}$, FC $3 m.s^{-1}$). The comparison of Figures 2-3 indicates the following:

- FC: does not look like any known gait type.
- Gallop: alternates back and front legs but does not naturally induce the rotatory aspect of the gait.
- Trot: achieved to generate a real trot gait.
- Walk: achieved to generate a real walk gait.

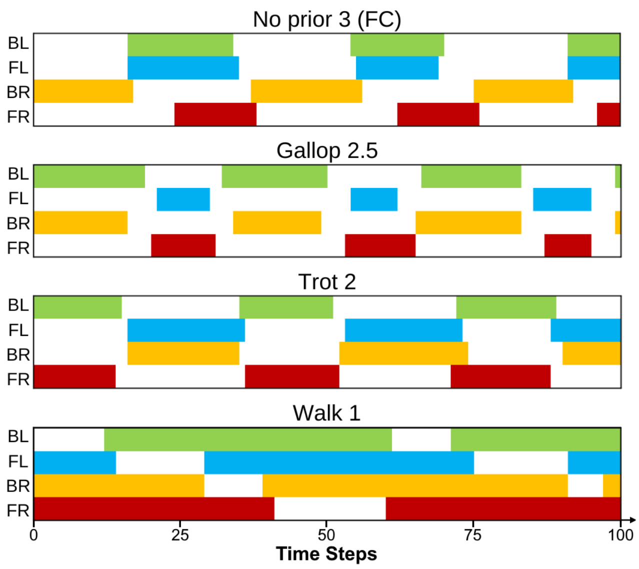

Figure 3: Step order of trained models.

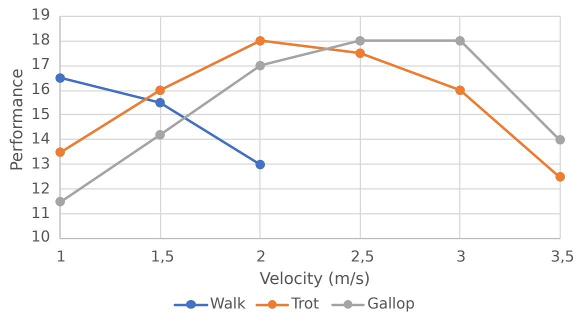

Figure 4: Performance index of trained models.

3.2 Energy performances

The performances of each gait type at each velocity is shown in Figure 4. As expected, the walk gait is the most effective in the lower velocities, the trot gait for medium velocities and the gallop gait for higher velocities.

4 Conclusion and Discussion

This paper proposed deep reinforcement learning for tailorable natural quadruped gait generation. This method involves new reward design by introducing reward related to the step order to learn the desired gait type, it successfully generates tailored different modes such as gallop, trot and walk. Moreover, generated gaits have similar properties in terms of energy efficiency with animal gaits, proving that even with a very simplistic model, DRL allows us to observe natural-like gaits and reproduce their key properties. Based on current results, future research should focus on introducing gait transitions for optimal velocity change and ensuring our method effectiveness in quadruped locomotion.

Acknowledgments

This work was supported by the JSPS Grant-in-Aid for Scientific Research on Innovative Areas "Hyper-Adaptability" project (Grant Number 22H04764).

References

[1] K. Kiguchi et al,"Energy-optimal gait analysis of quadruped robots," *Artificial Life and Robotics*, vol. 6, pp. 120–125, 09 2002.

[2] D. Owaki and A. Ishiguro, "A quadruped robot exhibiting spontaneous gait transitions from walking to trotting to galloping,"*Scientific Reports*, vol. 7, 12 2017.

[3] P. Arena et al, "Energy efficiency of a quadruped robot with neuro-inspired control in complex environments," *Energies*, vol.14, no. 2, 2021.

[4] J. Chai, D. Owaki and M. Hayashibe, "Deep Reinforcement Learning with Gait Mode Specification for Quadrupedal Trot-Gallop Energetic Analysis," Annu Int Conf *IEEE Eng Med Biol Soc*. 2021.

[5] E. Todorov et al, "Mujoco: A physics engine for model-based control," 2012 IEEE/RSJ *International Conference on Intelligent Robots and Systems*, pp. 5026–5033, 2012.

[6] T. Haarnoja et al, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018.

[7] Yoshua Bengio et al. "Curriculum learning," Proceedings of the 26th *Annual International Conference on Machine Learning*, 2009.

[8] B. Jean Carr, D. Dyrcus, "Canine gait analysis," *Today's Veterinary Practice 2016* Vol.6 No.2 pp.93-98, 2016.