2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Deep Reinforcement Learning with Gait Mode Specification for Quadrupedal Trot-Gallop Energetic Analysis

Jiazheng Chai, Dai Owaki and Mitsuhiro Hayashibe

Abstract— **Quadruped system is an animal-like model which has long been analyzed in terms of energy efficiency during its various gait locomotion. The generation of certain gait modes on these systems has been achieved by classical controllers which demand highly specific domain-knowledge and empirical parameter tuning. In this paper, we propose to use deep reinforcement learning (DRL) as an alternative approach to generate certain gait modes on quadrupeds, allowing potentially the same energetic analysis without the difficulty of designing an ad hoc controller. We show that by specifying a gait mode in the process of learning, it allows faster convergence of the learning process while at the same time imposing a certain gait type on the systems as opposed to the case without any gait specification. We demonstrate the advantages of using DRL as it can exploit automatically the physical condition of the robots such as the passive spring effect between the joints during the learning process, similar to the adaptation skills of an animal. The proposed system would provide a framework for quadrupedal trot-gallop energetic analysis for different body structures, body mass distributions and joint characteristics using DRL.**

I. INTRODUCTION

Quadruped robot is a common research area and there are numerous research topics which revolve around it, such as energetic studies [1][2][3], design principles [4], gait transition studies [5], etc. These studies play an important role in shedding light on the gait nature of quadrupeds under different circumstances such as varying walking speed [5] and terrain conditions [6], giving us a better understanding of quadrupeds as well as insights on better control strategies for quadrupeds. However, to carry out gait studies on quadruped robots, researchers have always been relying on hand-crafted controllers to generate various gait locomotion on a case-by-case basis [3][5][6][7][8], requiring domain expertise and time-consuming parameters tuning. Therefore, it is desirable to have a more general strategy which allows specific gait generations on quadruped robots with minimum fine-tuning in the search for optimal parameters.

In recent years, deep reinforcement learning (DRL) has been gaining attention as an alternative to classical controllers in quadruped robotic research [9][10][11]. This can be due to several advantages that DRL has over classical control strategies. One of the advantages is that the robotic

This work was supported by the JSPS Grant-in-Aid for Scientific Research on Innovative Areas Hyper-Adaptability Project under Grant 20H05458.

J. Chai, D. Owaki and M. Hayashibe are with the Neuro-Robotics Lab, Department of Robotics, Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan chai.jiazheng.q1@dc.tohoku.ac.jp owaki@tohoku.ac.jp hayashibe@tohoku.ac.jp

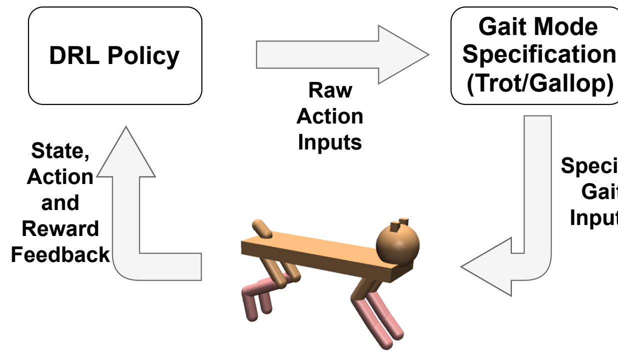

Fig. 1. The control loop of a quadruped using deep reinforcement learning (DRL). The raw action inputs are transformed into specific gait inputs as specified by the user in the learning process, allowing the production of the desired gait mode locomotion.

agents trained with DRL has the ability to generalize over various situations unseen during training [12][13], giving the robots adaptation skills similar to animals. DRL also requires less parameter tuning, providing that the reward function is well designed, as the learning process will find a set of optimal parameters via the optimization process.

Unfortunately, DRL has some disadvantages. It is well known that DRL requires long training time [14] and it increases with the complexity of the robots. In the case of quadruped research, there is no guarantee that the DRL-trained agents will finally possess a well-known gait type, making it difficult to carry out the same analysis as in the case of classical controllers. There is however a handful of work such as [15] which imposes a gait type or mode on a quadruped system by introducing prior knowledge in the DRL learning process.

In this paper, motivated by the potential of DRL in quadruped system studies, we propose some ways to overcome the downsides of DRL such that it is possible to be used as an alternative to classical controllers in quadruped energetic studies. Some of the contributions of this paper are:

- A method to specify a certain gait mode in the DRL learning process, ensuring a deterministic gait type in the output.
- Demonstrate that the gait mode specification speeds up the learning process and allows energetic study between two gait modes, i.e. the gallop and trot gaits, for two different forward speeds.
- Demonstrate the advantage of DRL in exploiting the body condition of the robots, i.e. the passive joint-spring effect, similar to the adaptation skills of an animal. This

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

also shows that the optimal set of parameters will be determined using the same algorithm regardless of the physical condition of the robots, in contrary to the case of classical controllers.

II. Method

A. Simulated Quadruped Agent

In this paper, we will employ a quadruped model (Fig. 1) simulated using the Mujoco physics engine [16]. It is a famous engine used in vaious DRL research [17][18] as it simulates realistic physical properties, such as the interaction between the feet of the quadruped and the ground. As this paper is the first step to validate our idea, we are content with using a simulated agent as it is low-cost and allows fast experimentation. The agent has four limbs, with each limb having three joints, i.e. the hip joint, the knee joint and the foot joint. The stiffness and damping parameters of the joints can also be modified to simulate various degrees of passive joint-spring effect during the running motion.

B. Deep Reinforcement Learning

While it is possible to use any off-the-shelf DRL algorithms to carry out our experiments, we have chosen the state-of-the-art DRL algorithm, i.e. the SAC [17] algorithm. The trained DRL policy will represent a complex function which outputs the joint torques for the quadruped agent as a function of time. The key characteristic of DRL concerned in this paper is the reward function employed during the learning process. For our experiments, we require that the robot moves forward in a two dimensional plane at a certain speed while considering at the same time its energy consumption issue. This can be translated to the reward function defined by the Equation (1):

$$R(t) = -|v(t) - v_{target}| - 0.1 \cdot \sum_i A_i(t)^2 \qquad (1)$$

At each time step t, the algorithm has to minimize two terms in the reward function. The first term requires that the current speed of the quadruped robot, $v(t)$ matches as close as possible to a given target speed v_{target}. The second term of the reward function is a representation of the energy consumed by the quadruped agent, where $A_i(t)$ is the magnitude of the torque input for the joint i. This term is scaled by a small coefficient so that the algorithm will not converge to a sub-optimal solution of not moving at all.

In the following of this paper, the performance metric is the distance travelled by the quadruped robot during a simulation of total time steps T, with each time step Δt being evaluated as one unit of time in the simulation environment. The performance metric can be written as the Equation (2):

$$Performance = \sum_T v(t) \cdot \Delta t \qquad (2)$$

In this study, we have used the energy expenditure index,which is the sum of the second terms of the reward function (1) throughout a running trajectory, which can be written as the Equation (3):

$$Energy\ index = \sum_T \sum_i A_i(t)^2 \qquad (3)$$

C. Gait Mode Specification

To specify a gait mode in the DRL learning process, we will exploit the information that we have about certain gait modes. In this paper, we are only considering the gallop and the trot gaits. For the gallop gait, we know that the limbs on the right side of the quadruped move symmetrically to the limbs on the left side. In reality, there might be a slight phase delay between the left and the right limbs in a gallop motion, but we make an assumption that there is no phase delay in this study. Therefore, the gallop gait is specified by having the DRL policy to produce torque inputs for the left limbs of the quadruped, then these torque inputs are copied identically for the right limbs, imposing the symmetric property in a gallop motion. This is translated as the Equation (4) where i and j are the corresponding joints on the right side and the left side of the quadruped respectively.

$$\tau_{right_i} = \tau_{left_j} \qquad (4)$$

It must be noted that this does not constrain the DRL algorithm from finding an optimal solution as it can still freely output the torque for the left limbs while receiving feedback about the overall body kinematic condition. For the trot gait, the limbs on the right and the left side move asymmetrically to each other. Therefore, the DRL policy torque inputs for the left limbs are negated and copied to the right limbs to impose a trot gait, as described by the Equation (5):

$$\tau_{right_i} = -\tau_{left_j} \qquad (5)$$

III. Experimental results

For all the experiments, the quadruped robot is trained for 400 thousand time steps until convergence. Three trials of each experiment are conducted and the average results as well as the standard deviations are presented. The video for the quadruped locomotion can be found at `https://youtu.be/RD4Uvskp9Zg`.

A. Gait Mode Specification Effects

The effect of the gait mode specification on the DRL learning process is presented in this subsection. As illustrated on the top of the Figure 2, the performance of the quadruped robot with the trot and gallop mode specified converged faster than the case without any specification, showing that the gait mode specification has indeed sped up the DRL learning process. On the bottom of the Figure 2, we can notice that the energy consumption for all cases peaked near the beginning of the learning process as the algorithm was exploring a gait locomotion starting from random movements. This corresponds to the beginning phase of the performance graph on the top of the Figure 2 where the performance increased steeply. The energy consumption decreased steadily in the remaining of the training process as the DRL algorithm discovered a more energy efficient

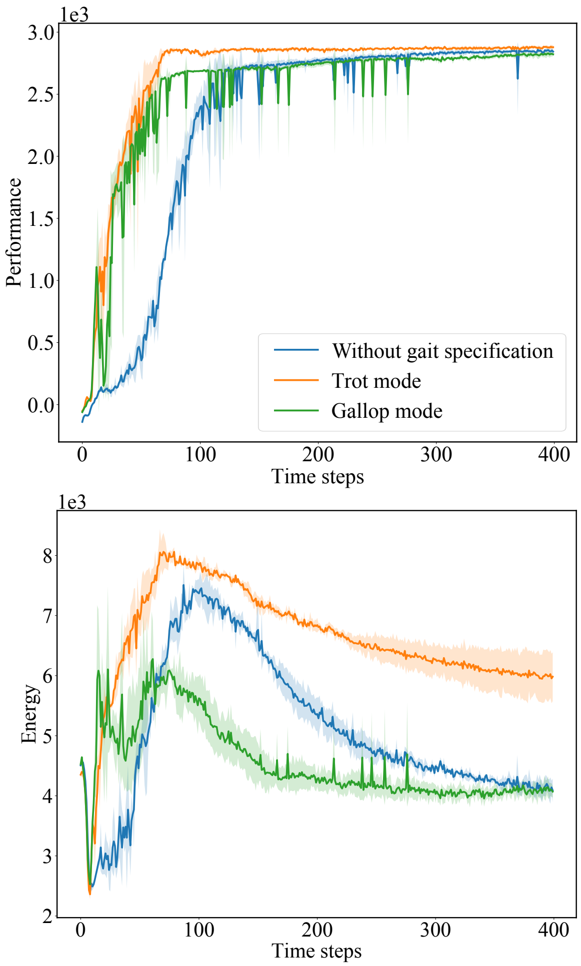

Fig. 2. The performance (top) and energy metric (bottom) comparison throughout the training process between DRL-trained agents with no gait specification (blue), with a trot mode specification (orange) and with a gallop mode specification (green).

locomotion to move forward while spending less energy at the same time, as specified by the second term in the reward function (1). This clearly demonstrates the advantage of using DRL over a classical controller as the reward function can be easily tailored to take into account different aspects when carrying out a task, much like the learning process of living things.

Besides the faster convergence of the DRL learning process, the gait mode specification also successfully imposed a predetermined gait type on the quadruped robot. As shown on the left of the Figure 3, the gait diagram of the quadruped robot without any gait specification does not correspond to any known gait type. It is a mix between the gallop gait and the trot gait. Indeed, there is no guarantee that the output locomotion of a DRL-trained robot would possess the desired gait type, rendering the analysis done using a classical controller [1][2][3][5] impossible. However, by specifying a certain gait mode in the DRL training process using our method, the output gait type resembles a well-known gait type, as shown by the gait diagram of the gallop mode and the trot mode in Figure 3.

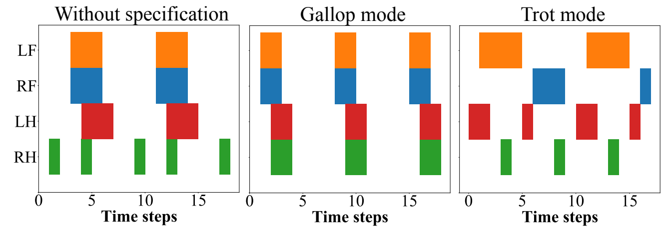

Fig. 3. The gait diagram for quadruped robots with different gait mode specification. LF and RF represent the left fore limb and the right fore limb respectively, while LH and RH represent the left hind limb and the right hind limb respectively. The colored regions correspond to the stance phase of each limb.

B. Energetic Study between Gallop and Trot Gaits

Studies such as [1][2][5][6] suggest that certain gait types are more suitable for quadruped robots moving at certain speed. In particular, for moving at a higher velocity, the gallop gait is shown to be more energy efficient while for a lower walking speed, the trot gait is believed to be preferable. Motivated by this result, we have conducted a performance and energetic analysis on the quadrupedal gait motions generated by the DRL algorithm for two different target speeds, i.e. a target speed of 3 $m/\Delta t$ and a target speed of 5 $m/\Delta t$, where Δt is one time step of simulation. From the performance graph on the top of the Figure 4, for moving forward at a speed of 3 $m/\Delta t$, the performance of the trot gait is slightly higher than the performance of the gallop gait. However, for the moving speed of 5 $m/\Delta t$, the gallop gait is better than the trot gait. While the performance difference is small, this is still an encouraging result and shares a similarity with the previously established results mentioned earlier. In term of the energy consumption during the forward motion, we can observe on the bottom of the Figure 4 that the energy consumed by the galloping quadruped robots is always significantly lower than the trotting quadruped robots for both moving speeds. While this does not meet our expectation that the trot gait would consume less energy at a lower forward speed, one reason could be that the forward speed of 3 $m/\Delta t$ is not slow enough and the trot gait is a running trot gait in reality. It could also be that the passive joint-spring parameters of our quadruped robotic model favors the gallop gait in our study. This is reasonable as studies such as [19] has shown that the gallop gait is more energy efficient as it exploits the passive spring energy stored between the joints, helping the quadruped robot to move forward easily. However, as the current study is still at the early stage of a more complete study, we found that the current result is promising as it shares some findings established in previous studies.

C. DRL Exploitation of the Passive Joint-Spring Effect

In order to verify that the passive joint-spring effect plays an important role in the energy efficiency of the forward running motion of a quadruped as stated in [19][20], we have varied the stiffness and the damping parameters of the joints of a galloping quadruped robot from the default parameters

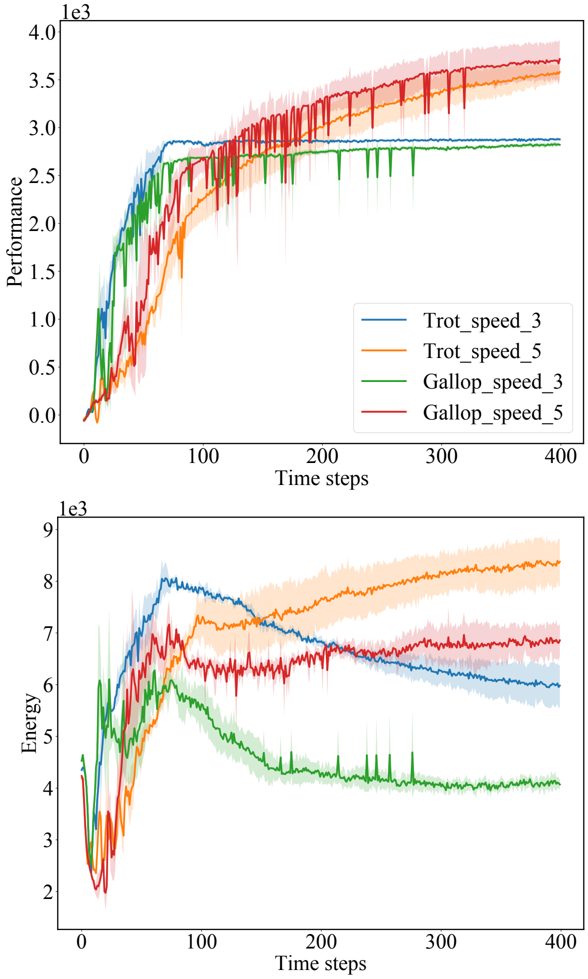

Fig. 4. The performance (top) and energy metric (bottom) comparison between DRL-trained agents with a trot mode specification and a target speed of 3 (blue); with a trot mode specification and a target speed of 5 (orange); with a gallop mode specification and a target speed of 3 (green); with a gallop mode specification and a target speed of 5 (red). The random spikes on the curves are the deviations occurred during the DRL learning process.

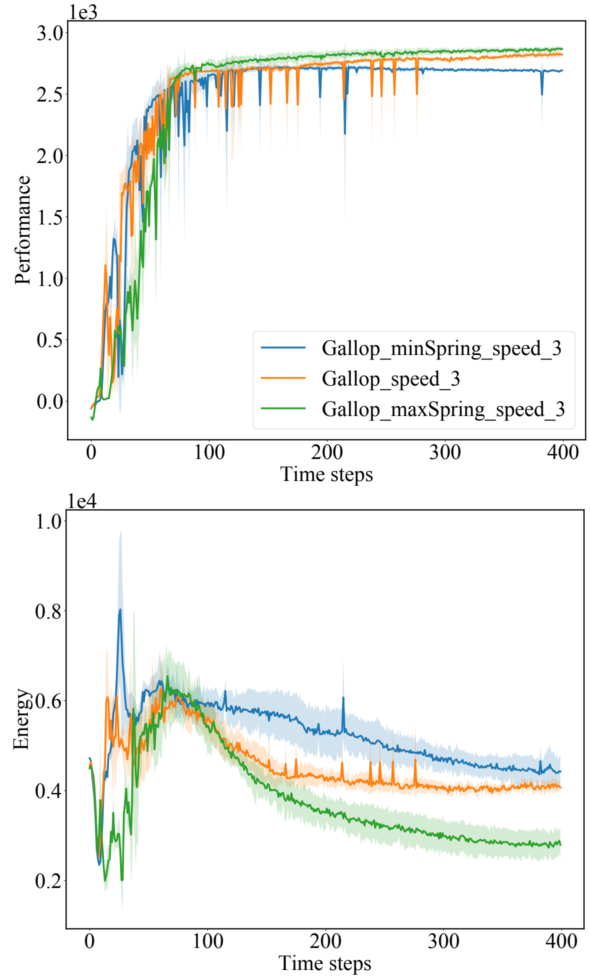

Fig. 5. The performance (top) and energy metric (bottom) comparison between DRL-trained agents with varying joint-spring effects, i.e. the minimum joint-spring effect during running (blue), the default joint-spring effect (orange), and the maximum joint-spring effect during running (green). All the agents are with a gallop mode specification and a target speed of 3.

used in the previous sections. To have a more joint-spring effect, we decreased the stiffness and the damping parameters for all the joints. On the other hand, to have a less joint-spring effect, the stiffness and the damping parameters are increased. As illustrated by the performance graph on the top of the Figure 5, the experimental results clearly show that the more joint-spring effect a quadruped robot has, the higher the galloping performance of the robot. In addition, from the energy curves on the bottom of the Figure 5, the energy consumption of the galloping motion also decreases as the joint-spring effect increases. This result matches perfectly with [19][20] as the passive joint-spring aids the running motion and reduces the energy consumed by the quadruped robot to gallop forward. This result also supports the idea that the DRL algorithm can serve as a general algorithm that can adapt to different physical conditions of a quadruped robot to produce gait motions for analysis, contrary to the case of classical controllers. Indeed, in the case of using classical controllers to generate gait motions, the parameters of the controllers need to be hand-tuned whenever there are changes in the physical properties of the quadruped robot or the experimental environment. DRL algorithms clearly have an advantage over classical controllers in this regard [12][13].

IV. Discussion

From our experimental results, by using the proposed gait mode specification method, we have successfully imposed a certain gait mode on the quadruped agent, eliminating the random gait mode problem of a standard DRL algorithm. At the same time, the proposed method has equally sped up the learning convergence speed of the DRL algorithm. We have also demonstrated that DRL algorithms are promising as an alternative to classical controllers for gait generations in quadruped energetic analysis. Coherent results were obtained between our experiments and some related previous works [1][2][5][6]. The advantage of DRL in exploiting the dynamics properties of the quadruped agent is also demonstrated through the passive joint spring experiment.

While the current results are promising, however, there are also several points that could be improved in the future. Currently, the introduced gait mode specification works as our expectation to impose a certain gait type on the DRL-trained quadruped robot. We believe that a more complicated gait mode specification method could be introduced to output a more precise gait locomotion. For example, as we have mentioned earlier that for a gallop gait, there might be a slight phase delay between the limbs on each side of the quadruped. A more complete gait mode specification method could possibly deal with this issue.

In the future, a more realistic quadruped model could be used for experiments. Ultimately, a real quadruped robot could be employed to carry out the same experiments in this paper to verify that the DRL-produced gait motion is plausible as well in a real robot. Ideally, we would also like to experiment with different gait types, so that we can repeat the analysis done in works such as [1][2][5][6] using DRL algorithms instead of classical controllers for gait generations.

V. CONCLUSIONS

In this paper, we have proposed the gait mode specification method which speeds up the convergence of a DRL algorithm as well as imposing a certain gait type on the quadruped robot, contrary to the case without any specification. DRL algorithms have also been demonstrated to be potentially an alternative to classical controllers for quadruped gait generations for energetic analysis. Moreover, DRL algorithms equally show the ability to generalize to situations never seen before, providing optimal locomotion and removing the need for tedious manual parameters tuning in classical controllers, as demonstrated in the passive joint spring exploitation experiment. We believe that our work is the first step towards a more general framework of locomotion analysis in quadrupeds using DRL, contributing to the research field of understanding quadrupedal motion control on gait coordination.

REFERENCES

[1] K. Kiguchi, Y. Kusumoto, K. Watanabe, K. Izumi, and T. Fukuda, "Energy-optimal gait analysis of quadruped robots," *Artificial Life and Robotics*, vol. 6, pp. 120–125, 09 2002.

[2] M. Silva and J. Tenreiro Machado, *Energy Efficiency of Quadruped Gaits*. Springer Berlin Heidelberg, 02 2006, p. 735–742.

[3] K. Yang, X. Rong, L. Zhou, and Y. Li, "Modeling and analysis on energy consumption of hydraulic quadruped robot for optimal trot motion control," *Applied Sciences*, vol. 9, p. 1771, 04 2019.

[4] S. Seok, A. Wang, M. Y. Chuah, D. Otten, J. Lang, and S. Kim, "Design principles for highly efficient quadrupeds and implementation on the MIT cheetah robot," in *2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, May 6-10, 2013*. IEEE, 2013, pp. 3307–3312. [Online]. Available: https://doi.org/10.1109/ICRA.2013.6631038

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

[6] P. Arena, L. Patanè, and S. Taffara, "Energy efficiency of a quadruped robot with neuro-inspired control in complex environments," *Energies*, vol. 14, no. 2, 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/2/433

[7] H. Chai, X. Rong, X. Tang, and Y. Li, "Gait-based quadruped robot planar hopping control with energy planning," *International Journal of Advanced Robotic Systems*, vol. 13, no. 1, p. 20, 2016. [Online]. Available: https://doi.org/10.5772/62140

[8] Y. Fukuoka, Y. Habu, and T. Fukui, "A simple rule for quadrupedal gait generation determined by leg loading feedback: A modeling study," *Scientific reports*, vol. 5, p. 8169, 02 2015.

[9] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning Quadrupedal Locomotion over Challenging Terrain," *arXiv e-prints*, p. arXiv:2010.11251, Oct. 2020.

[10] V. Tsounis, M. Alge, J. Lee, F. Farshidian, and M. Hutter, "Deepgait: Planning and control of quadrupedal gaits using deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–1, 03 2020.

[11] J. Chai and M. Hayashibe, "Motor synergy development in high-performing deep reinforcement learning algorithms," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1271–1278, 2020.

[12] Honglak Lee, Yirong Shen, Chih-Han Yu, G. Singh, and A. Y. Ng, "Quadruped robot obstacle negotiation via reinforcement learning," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, 2006, pp. 3003–3010.

[13] G. Bellegarda and Q. Nguyen, "Robust Quadruped Jumping via Deep Reinforcement Learning," *arXiv e-prints*, p. arXiv:2011.07089, Nov. 2020.

[14] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[15] B. Hu, S. Shao, Z. Cao, Q. Xiao, Q. Li, and C. Ma, "Learning a faster locomotion gait for a quadruped robot with model-free deep reinforcement learning," in *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2019, pp. 1097–1102.

[16] 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*, 2012, pp. 5026–5033.

[17] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018.

[18] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," 2018.

[19] N. C. Holt, T. J. Roberts, and G. N. Askew, "The energetic benefits of tendon springs in running: is the reduction of muscle work important?" *Journal of Experimental Biology*, vol. 217, no. 24, pp. 4365–4371, 2014. [Online]. Available: https://jeb.biologists.org/content/217/24/4365

[20] A. Sutrisno and D. J. Braun, "How to run 50% faster without external energy," *Science Advances*, vol. 6, no. 13, 2020. [Online]. Available: https://advances.sciencemag.org/content/6/13/eaay1950