LAGEPP – UMR 5007
Université Claude Bernard Lyon 1, bât 308G ESCPE
43 bd du 11 Novembre 1918, 69622 Villeurbanne Cedex

Laboratoire
d'automatique,
de génie des procédés
et de génie pharmaceutique

Research Internship:
« Learning robust policies for walking humanoids robots »

To apply : Submit your CV and a motivation letter to: vincent.andrieu@univ-lyon1.fr

Location: UCBL1 http://www.univ-lyon1.fr/
Lab: LAGEPP http://www.lagepp.univ-lyon1.fr

Advisors of the stage: Daniele Astolfi (CNRS LAGEPP), Vincent Andrieu (CNRS LAGEPP), Samuele Zoboli (LAGEPP)

Scientific Domain: Reinforcement learning, Control theory, mathematics, dynamical system, engineering.

Objectives, scientific challenges and expected original contributions:

In recent years, the use of techniques derived from artificial intelligence enabled major advances in the field of image processing and segmentation, see e.g. [6]. In the context of automated decision-making processes, deep neural networks-based techniques such as deep reinforcement learning [1,2] led to amazing results, see e.g., [3,4]. Nevertheless, these approaches provide more questionable results in the field of control of real-time systems. Such a discrepancy may be explained by the difficulty faced by these algorithms in generalizing over previously unseen data. Once applied to real-world scenarios, the learned policies may fail as they face the so-called *reality gap*. This is caused by inevitable differences between the simulated environment they were trained on, and the real physical system. In order to enrich the training dataset and improve robustness, standard methods propose to learn in parallel over a wide range of slightly different environments or in an adversarial context [5]. However, strong guarantees are still lacking.

In control theory, generalization is called robustness. Some known control methods ensure stability properties coupled with some kind of robustness guarantees.

The objective of this internship is to translate techniques from robust control, and more precisely from regulation theory, in the context of deep reinforcement learning. The idea is to add an internal model and learn a policy based on an enhanced reward incorporating these extra states. This will enable us to learn robust control policies. The proposed methodology will be tested on the challenging model of a humanoid robot learning to stand up, walk or run.

Internship organization: In the first part of the internship, the student will read and develop a new theoretical tool to design a reinforcement learning strategy which incorporates a prescribed internal model. In a second step, the MuJoco [7] numerical humanoid will be employed to show the effectiveness of the new control policy strategy.

Length and remuneration of the stage: 5 to 6 months with a salary of 550€ net per month, to be effectuated between January 2023 and September 2023.

Application and expected profile: We look for a candidate with a strong background in math, control theory or programming. A PhD grant on own funding (ANR) is available.

References :

[1] Sutton, Richard S., and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
[2] François-Lavet, Vincent, et al. "An introduction to deep reinforcement learning." *Foundations and Trends® in Machine Learning* 11.3-4 (2018): 219-354.
[3] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." *arXiv preprint arXiv:1312.5602* (2013).
[4] Lillicrap, Timothy P., et al. "Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971 (2015).
[5] Pinto, Lerrel, et al. "Robust adversarial reinforcement learning." *International Conference on Machine Learning*. PMLR, 2017.
[6] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
[7] https://github.com/deepmind/mujoco