
On-sky results with a real-time model-free reinforcement learning method

Bartomeu Pou^{*1,2}, Sergi Albiach¹, Vincent Deo³, Kyohoon Ahn³, Sebastien Vievard³, Julien Lozi³, Olivier Guyon^{3,4,5,6}, Eduardo Quinones¹, Mario Martin², and Damien Gratadour⁷

¹Barcelona Supercomputing Center (BSC) – Spain

²Universitat Politècnica de Catalunya (UPC) – Spain

³Subaru Telescope – United States

⁴Steward Observatory – United States

⁵Wyant College of Optical Sciences [University of Arizona] – United States

⁶Astrobiology Center of NINS – Japan

⁷Observatoire de Paris – LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Univ. Paris Diderot, Sorbonne Paris Cité, 5 place Jules Janssen, 92195 Meudon, France. – France

Abstract

This paper introduces a novel real-time non-linear predictive control approach based on a model-free reinforcement learning (RL) method, with a non-linear reconstruction relying on supervised learning (SL) using a U-net architecture. We present our on-sky and bench results obtained on the Subaru Coronagraphic Extreme Adaptive Optics (SCEXAO) instrument and show that our approach outperforms the standard modal integrator controllers. To develop our two-component model, we quantify the possible improvement of a non-linear reconstruction for the pyramid wavefront sensor (PyWFS) and a wavefront prediction depending on the intrinsic delay of the system.

This analysis provides the basis for building our two-component model of non-linear prediction with RL and reconstruction with SL. The RL component is trained online with telemetry data and the SL model is trained offline with a previously built dataset of WFS images and phases.

The proposed model involves training (RL) and inferring (RL and SL) at high framerates (2kHz PyWFS images). To run at such speed, we propose a scheme that we integrate into the Compute and Control for Adaptive Optics real-time control package (CACAO), leveraging this framework to process the data and both infer and train at speeds that allow for increased performance. The main ideas of the scheme are TensorRT for inference, multiple GPU training for the RL model and a high degree of parallelization.

Finally, we present our results demonstrating that our model can run at the required speeds and outperform a default modal integrator controller both on-sky and on the bench.

*Speaker

Keywords: Reinforcement Learning, High Performance Computing, Machine Learning, Real, time Control