Predictive RL for Continuous Trajectory Tracking

Georg Schäfer[1,2] Jakob Rehrl[1] Stefan Huber[1] Simon Hirlaender[2]

[1] Josef Ressel Centre for Intelligent and Secure Industrial Automation,
Salzburg University of Applied Sciences, Salzburg, Austria

[2] Paris Lodron University of Salzburg, Salzburg, Austria

1. The Industrial Problem

- Industrial cyber-physical systems require **smooth, highly precise control**.
- Standard deep reinforcement learning agents are often **purely reactive**, leading to **lag and overshoots** when tracking targets.
- Controllers like Model Predictive Control (MPC) **anticipate future states** but demand high online computational resources.
- **Our Goal:** Combine the **foresight of MPC** with the **fast inference** of deep reinforcement learning.

2. Methodology & Experimental Setup

- **Proximal Policy Optimization** applied to a **1-DOF Quanser Aero 2** testbed, representing non-linear industrial dynamics.
- Reactive Baseline State: $s_t = (\theta_t, \dot{\theta}_t, r_t)$
- **Predictive State:** $s_t = (\theta_t, \dot{\theta}_t, r_t, \dot{r}_t, r_{t+1 \cdot \Delta t}, ..., r_{t+N \cdot \Delta t})$
- with θ: pitch, $\dot{\theta}$: vel., r: target, $\dot{r}$: target vel., N: # future targets, Δt: step interval.
- **Evaluation:** 10 simulation runs per configuration evaluating absolute mean deviation on a fixed trajectory, followed by **zero-shot sim-to-real transfer** to the physical hardware.

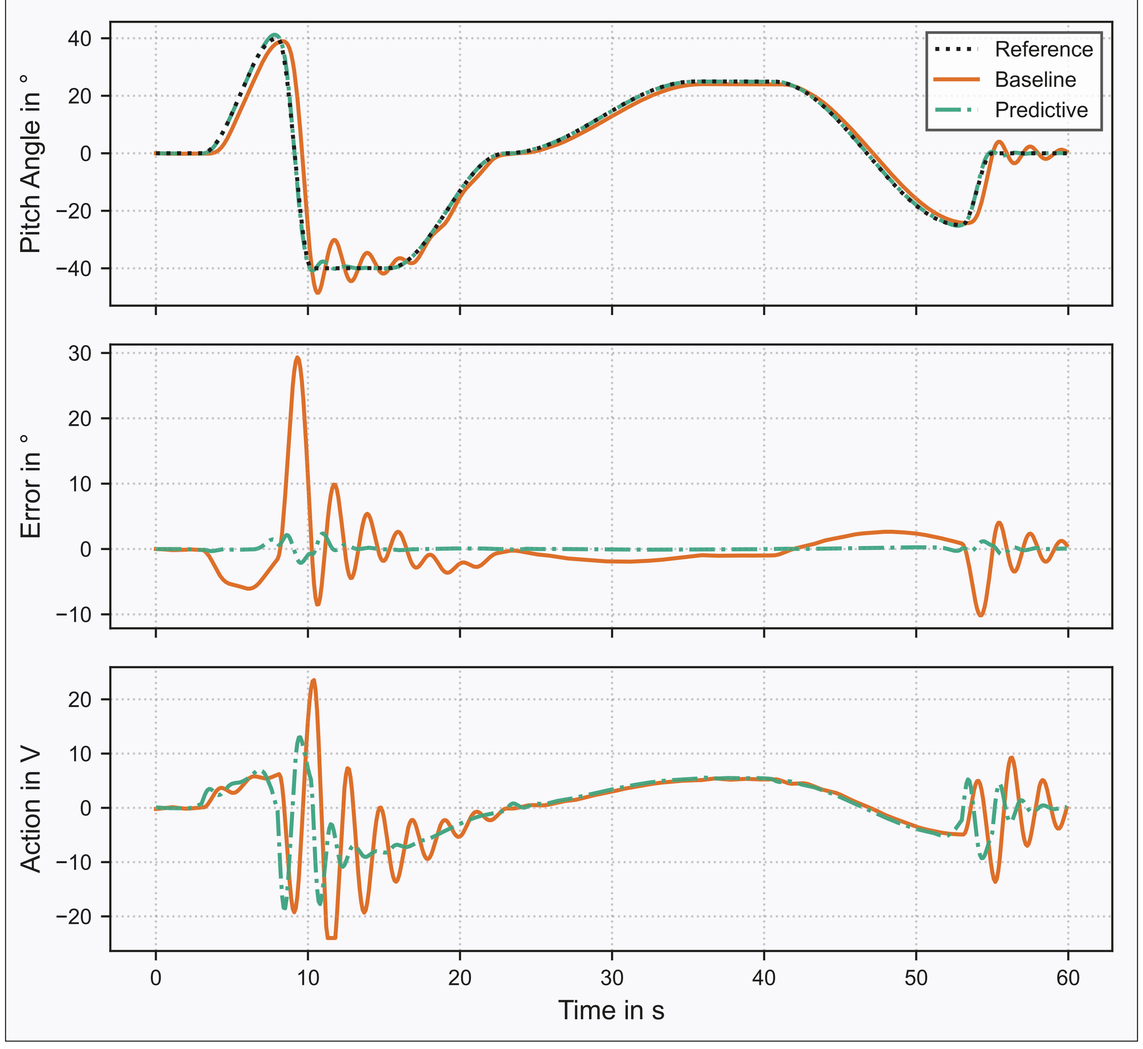
3. Results: Mean Absolute Deviation (°)

$\dot{r}$	N	Δt	Worst	Best	Mean	Std	Hardware
No	0	–	3.05	2.51	2.73	0.18	2.34
No	1	0.1	1.88	1.49	1.70	0.13	2.09
No	1	1.0	0.95	0.78	0.85	0.06	**1.11**
No	4	0.5	1.02	0.49	0.70	0.14	1.21
Yes	0	–	0.90	0.41	0.52	0.15	1.23
Yes	1	0.1	0.74	0.34	0.48	0.10	1.43
Yes	1	1.0	0.50	0.22	0.34	0.08	1.14
Yes	4	0.5	0.53	0.19	**0.31**	0.10	**1.11**

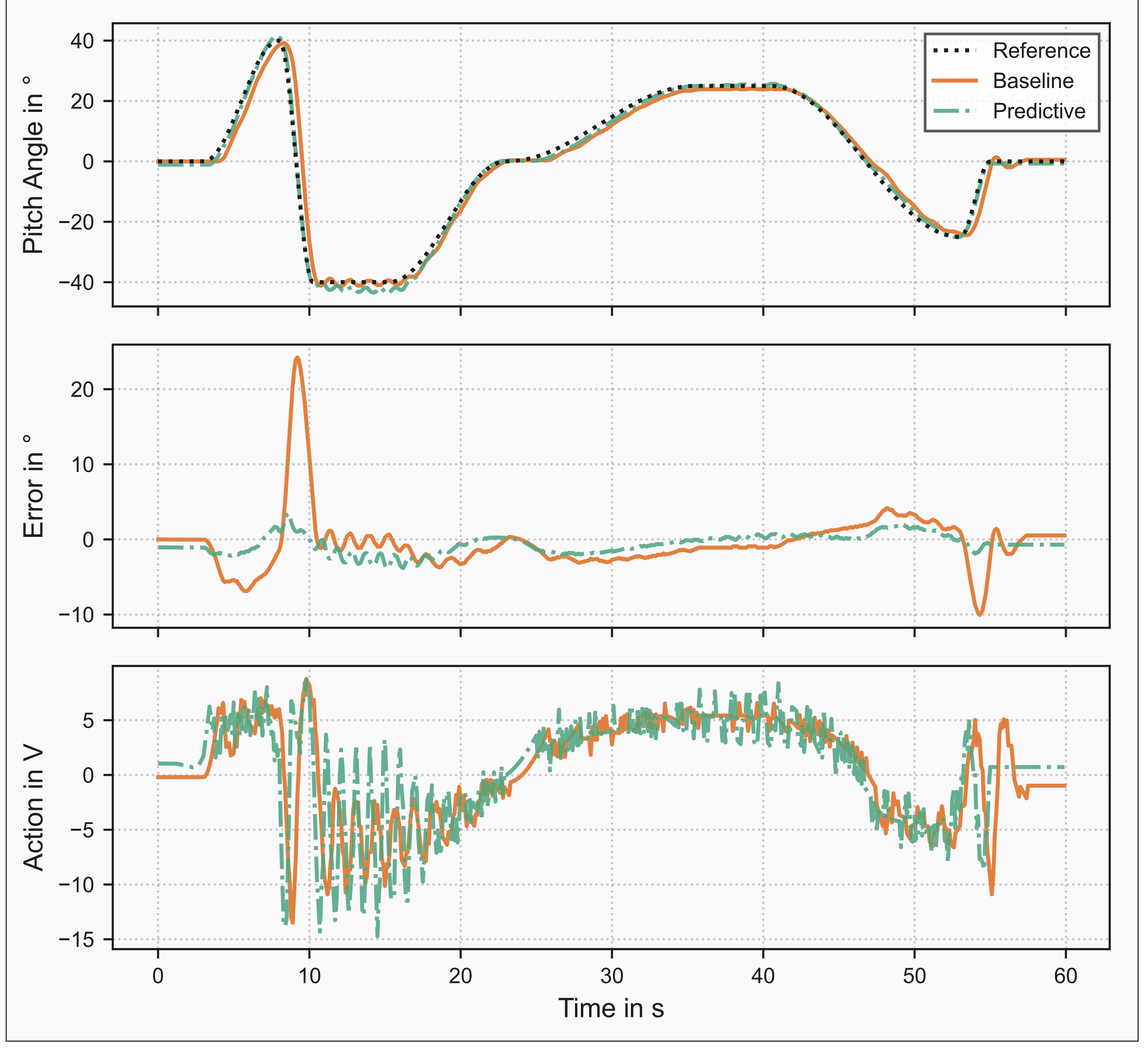
4. Conclusion & Outlook

- Adding a **1.0s prediction horizon** cut real-world tracking error in half (from **2.34° to 1.11°**).
- A **single future target** performed equally well as the most **dense predictive state**. This simplicity acts as a **regularizer**, achieving peak performance without needing velocity information.
- **Outlook:** Training on the physical hardware and benchmarking its inference speed and energy efficiency directly against MPC.

Simulation (Baseline vs. $\dot{r}$, N = 4, Δt = 0.5s)



Physical System (Baseline vs. $\dot{r}$, N = 4, Δt = 0.5s)



Interreg Bayern-Österreich

Kofinanziert von der Europäischen Union

FH Salzburg

PARIS LODRON UNIVERSITÄT SALZBURG
Lab for Intelligent Data Analytics Salzburg

RL4AA 2026 – Poster #46
Liverpool, United Kingdom
georg.schaefer@fh-salzburg.ac.at