Excluding the Irrelevant: Focusing Reinforcement Learning through Continuous Action Masking

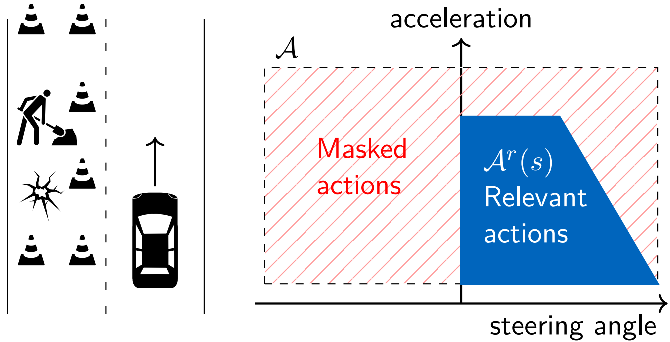
Roland Stolz[1], **Hanna Krasowski**[2],
Jakob Thumm[1], Michael Eichelbeck[1], Philipp Gassert[1], Matthias Althoff[1]

[1]Technical University of Munich, [2]University of California, Berkeley

The Thirty-Eighth Annual Conference on Neural Information Processing Systems

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1608 – 501798263, AL 1185/9-1, and AL 1185/33-1.

Leverage task knowledge to **constrain the action space** and **focus exploration** to relevant actions



Blocked lane segment **State-dependent masking**

Our **continuous action masking** concept

Concept

Encode domain knowledge in a state-dependent relevant action set $\mathcal{A}^r(s)$ to constrain the sampling from the policy to this set.

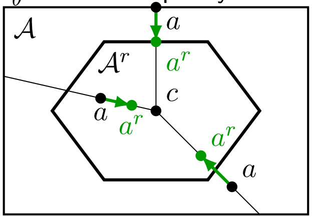
$$a^r \sim \pi_\theta^r(a^r|s) = h\big(\pi_\theta(a|s), \mathcal{A}^r(s)\big).$$

Assumptions

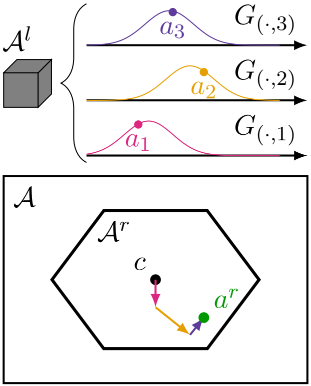
- The relevant action set $\mathcal{A}^r(s)$ is convex and can be computed in every state
- The policy is represented by a parameterized probability distribution $a \sim \pi_\theta(a|s)$

We propose **three approaches** for obtaining the **relevant policy**

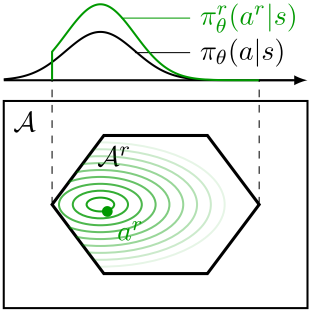
$\mathcal{A}^r$	relevant action set
$\mathcal{A}^l$	latent action set
a^r	relevant action
c	center of $\mathcal{A}^r$
G	Generator mat. of $\mathcal{A}^r$
π_θ^r	relevant policy



Ray mask



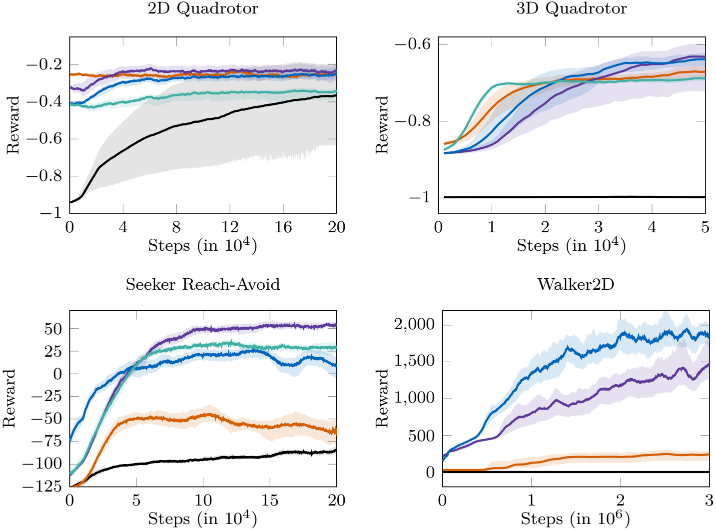
Generator mask



Distributional mask

The masking approaches improve **sample efficiency and performance** on four control environments

- Baseline PPO
- Replacement [1]
- Ray Mask
- Generator Mask
- Distributional Mask



[1] H. Krasowski et al., *Provably safe reinforcement learning: Conceptual analysis, survey, and benchmarking*, TMLR 2023

Focusing on state-dependent **relevant action sets** with continuous action masking

Key findings

By constraining the action space of the RL agent, continuous action masking can

- incorporate **domain knowledge**,
- improve **sample efficiency** and **convergence**,
- provide **safety guarantees**.

Future work

- deterministic policies
- non-convex and disjoint relevant action sets

Interested? Relevant to your application? Questions?

Please contact us!

Roland Stolz
roland.stolz@tum.de

Hanna Krasowski
krasowski@berkeley.edu

Jakob Thumm

Michael Eichelbeck

Philipp Gassert

Matthias Althoff