[1] Fraunhofer IKS
[2] SIEMENS

Felippe S. Roza[1], Simon Hadwiger[2], Ingo Thorn[2], Karsten Roscher[1]

Towards Safety Assurance of Uncertainty-Aware Reinforcement Learning Agents

AI/RL capabilities

Motivation

© Fraunhofer IKS

RL is not the choice for safety-critical applications

Motivation

- Industrial players are historically very conservative
- High complexity of DNNs
- Standard verification and validation are not compatible to Deep RL
- Unreliable under distributional shifts

© Fraunhofer IKS
Public information

Contribution

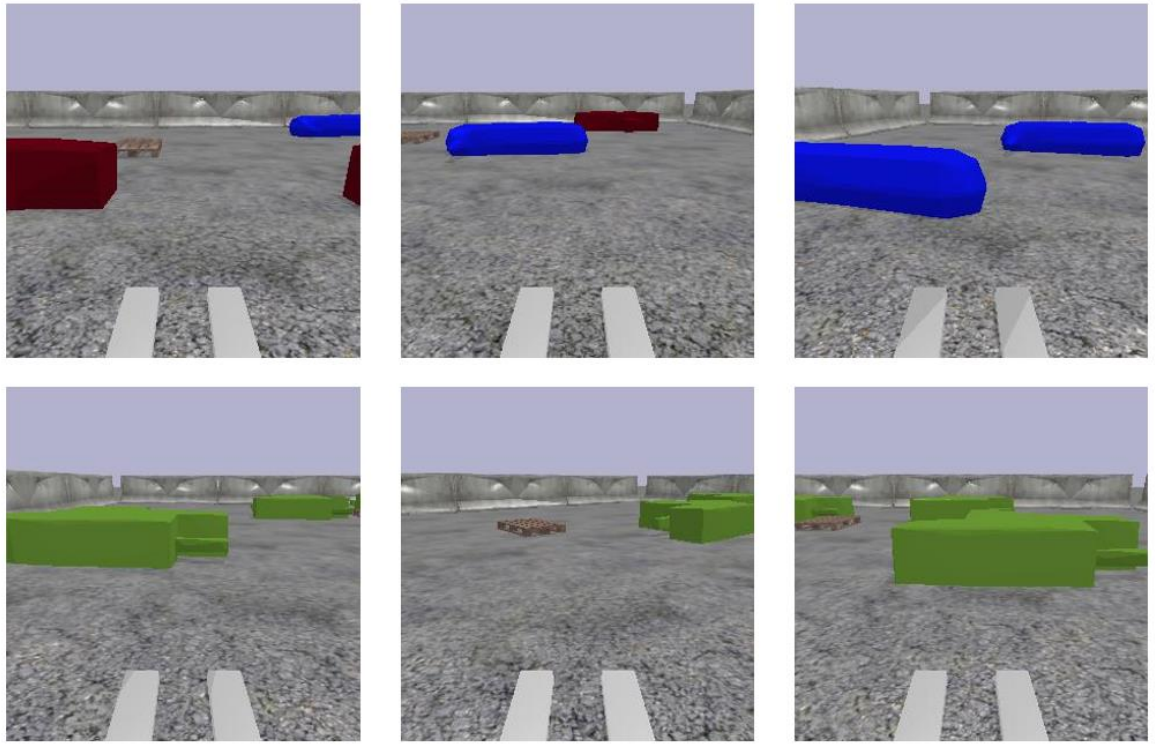
As a position paper:

1. **Propose an architecture that integrates RL and uncertainty estimation**
2. **Provide evidence that uncertainty-based OOD detection can help identifying unsafe states**
3. **Outline future steps towards building a safety case for uncertainty-aware RL systems**

© Fraunhofer IKS

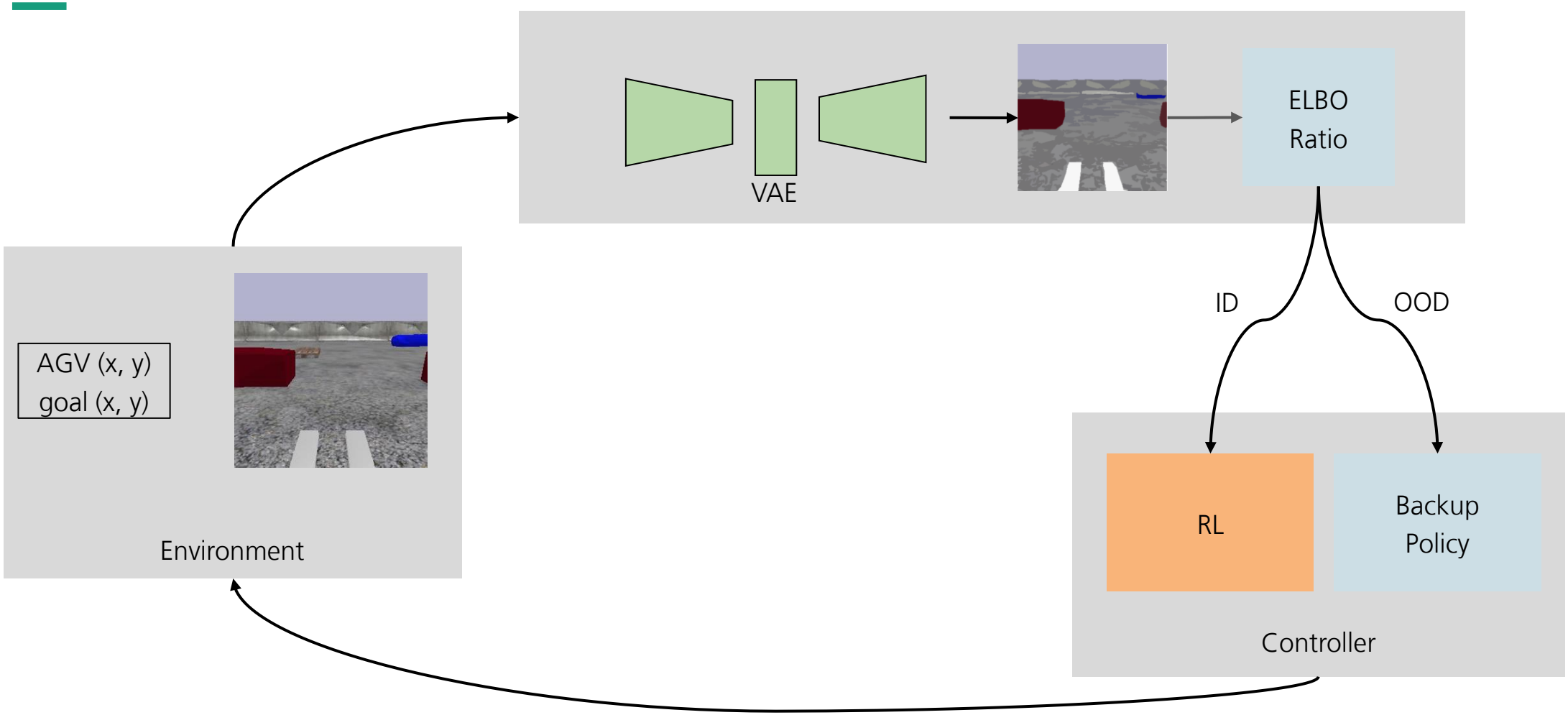
Use case

- Moving goods in a warehouse
- Goal: reach the pallet while avoiding obstacles
- Custom AGV/forklift environment
- ID vs OOD obstacles: different colors and shapes



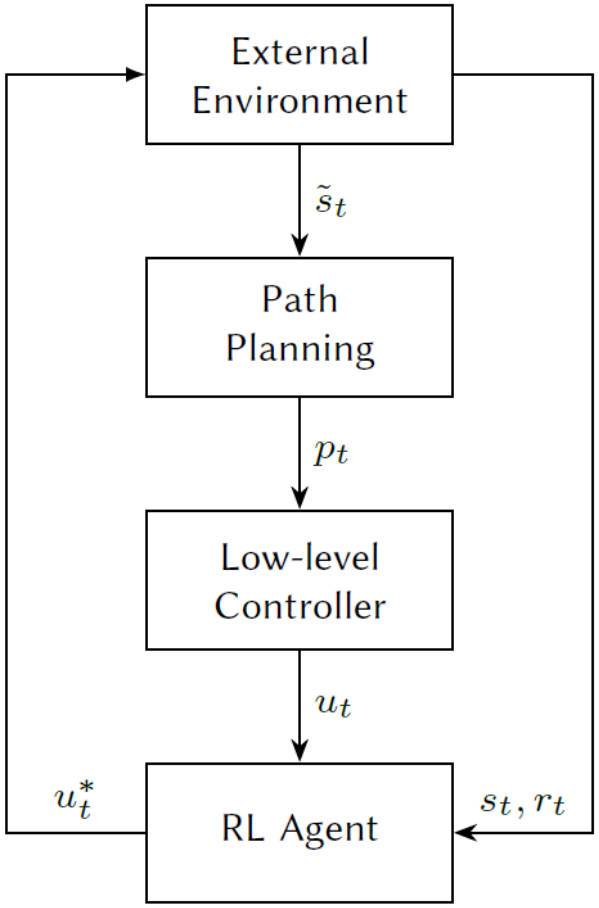
© Fraunhofer IKS
Public information

Uncertainty-aware RL



© Fraunhofer IKS
Public information

RL-based controller



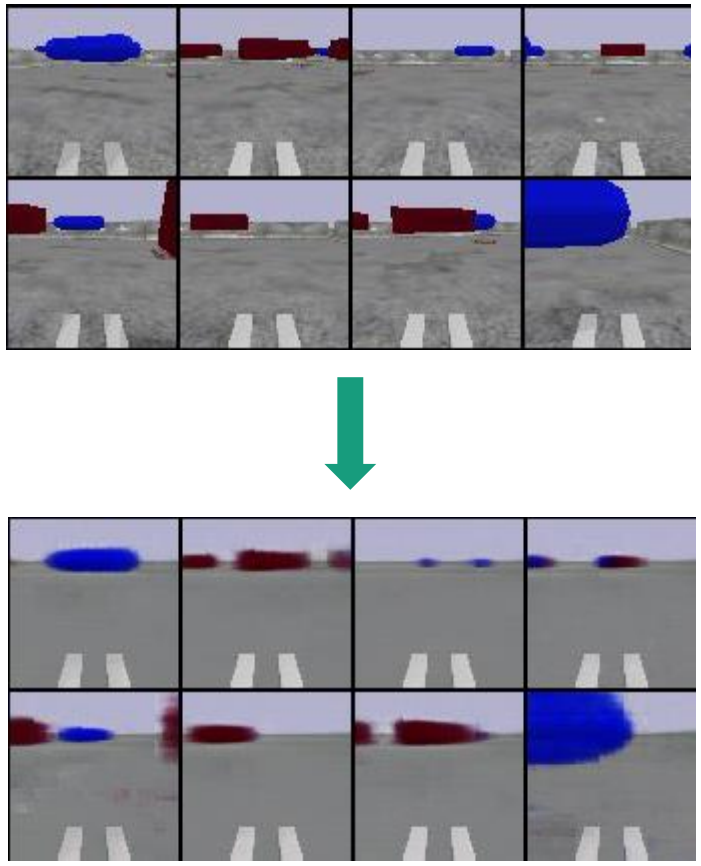
- Path planner calculates optimal trajectory
- Low-level controller calculates control actions
- RL - obstacle avoidance
- Better fit for industrial-grade applications

© Fraunhofer IKS
Public information

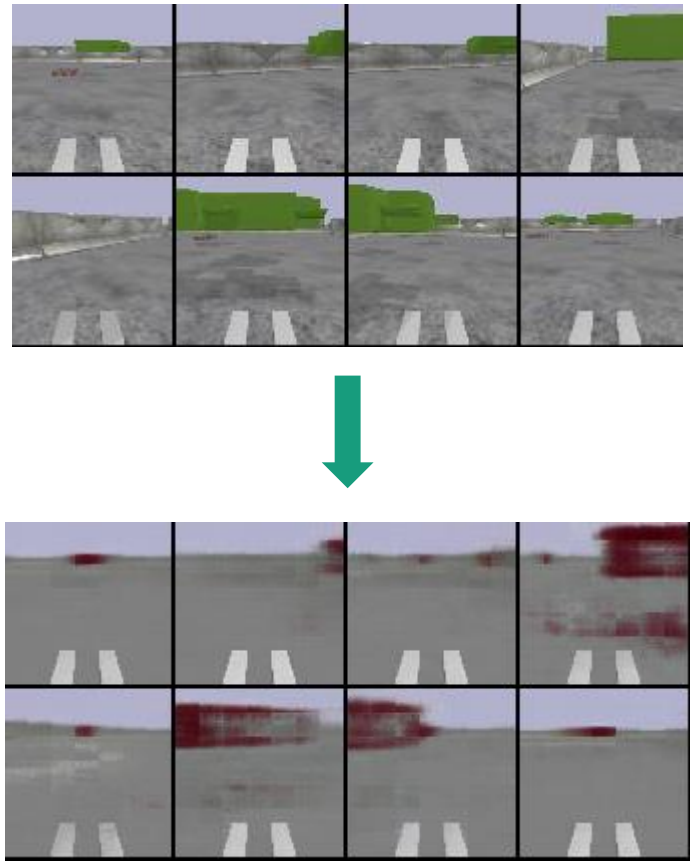
VAE reconstruction

Preliminary results

ID



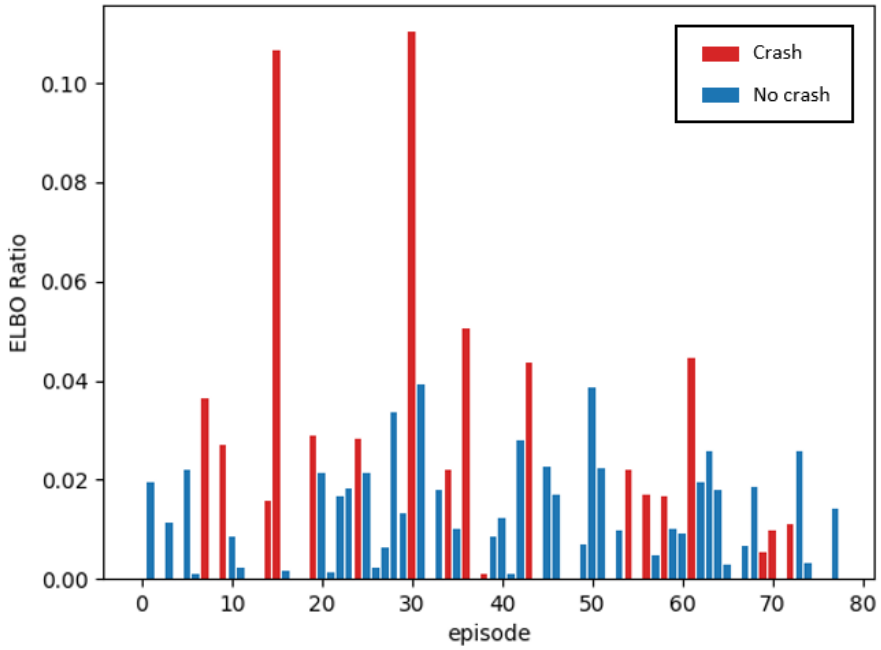
OOD



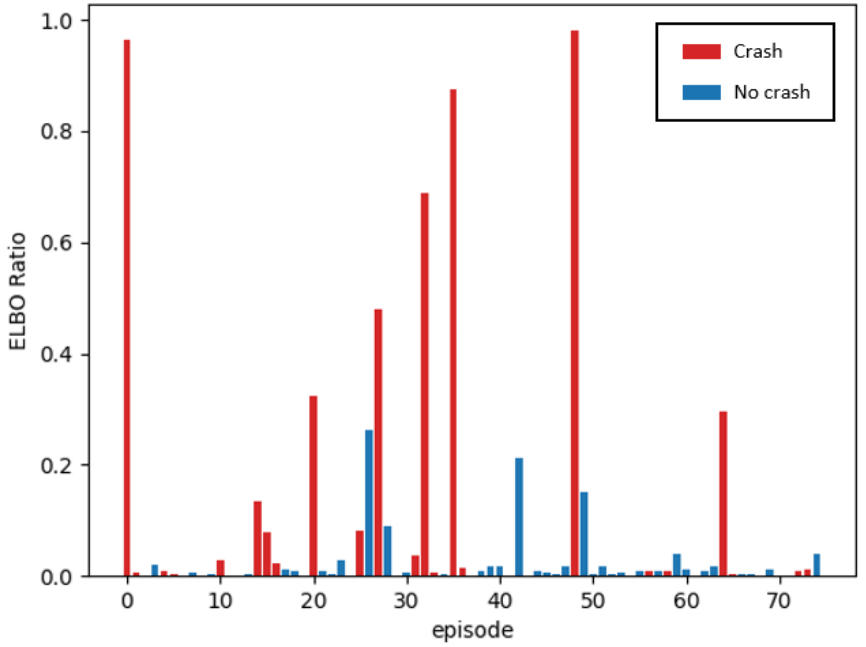
© Fraunhofer IKS
Public information

Uncertainty scores in termination states
Preliminary results

ID



OOD



© Fraunhofer IKS
Public information

Future perspective

Shortcomings

- Only covers a single aspect of safety in RL
- Limited testing scenarios
- Naïve/simplistic OOD scenarios
- Limitations of VAE as OOD detector

Future steps

- Extensive experimentation
- Reason behind FPs and FNs
- Impact of OODD’s residual error
- Failure rate calibration
- SOTIF

© Fraunhofer IKS
Public information

We'd love to hear from you!

Felippe Schmoeller da Roza
felippe.schmoeller.da.roza@iks.fraunhofer.de

https://iks.fraunhofer.de
https://safe-intelligence.fraunhofer.de

RL reward function

Encourage following optimal trajectory:

$r = 0.0 \; if \; u_t{}^* \cong u_t$
$r = -0.1$ otherwise

Accomplish task

$r = 100$ if reaching the goal
$r = -10$ if timing out

Obstacle avoidance

$r = -100$ if crashing

© Fraunhofer IKS
Public information