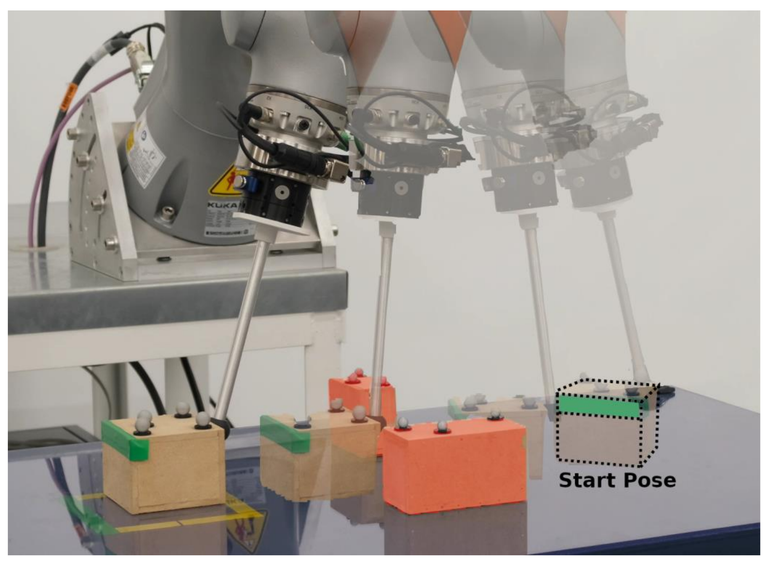
Learning Goal-Directed Object Pushing in Cluttered Scenes With Location-Based Attention

Nils Dengler*[1], Juan Del Aguila Ferrandis*[2], Joao Moura[2], Sethu Vijayakumar[2], and Maren Bennewitz[1]

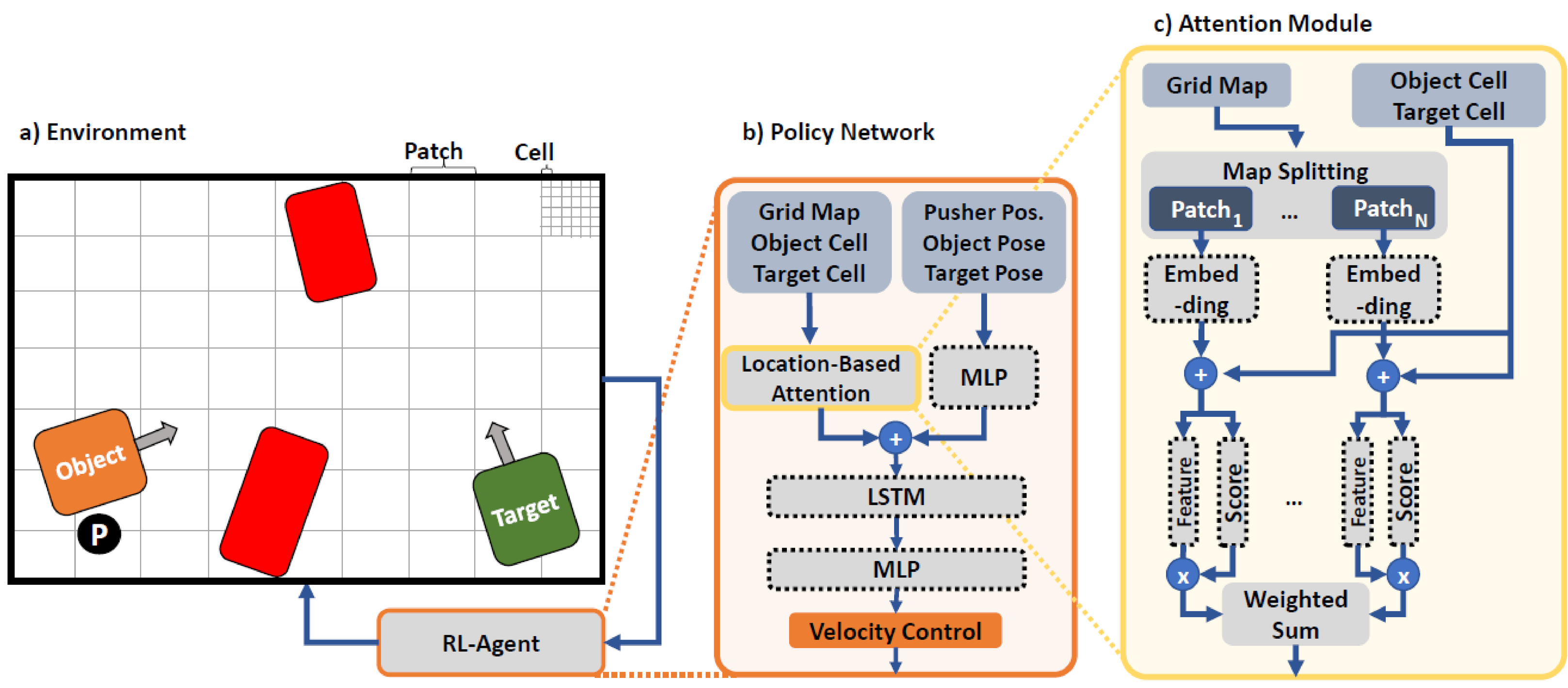
1. Humanoid Robots Lab, University of Bonn, Germany
2. School of Informatics, The University of Edinburgh, Edinburgh, UK

Motivation

- How to precisely push objects in cluttered scenes?
- Which state representation is needed?
- How can attention mechanisms improve the pushing?



Method Overview



Our Approach

- Model-free RL learning approach
- Location-based attention feature extraction for obstacle aware pushing behaviour
- Categorical exploration PPO to capture multimodal behaviour in planar pushing (slide, separating, stick)

Action: End-effector velocity (v_x, v_y)

Observation: Object and target poses (x, y, θ), EE position (x, y), binary occupancy grid

Reward:

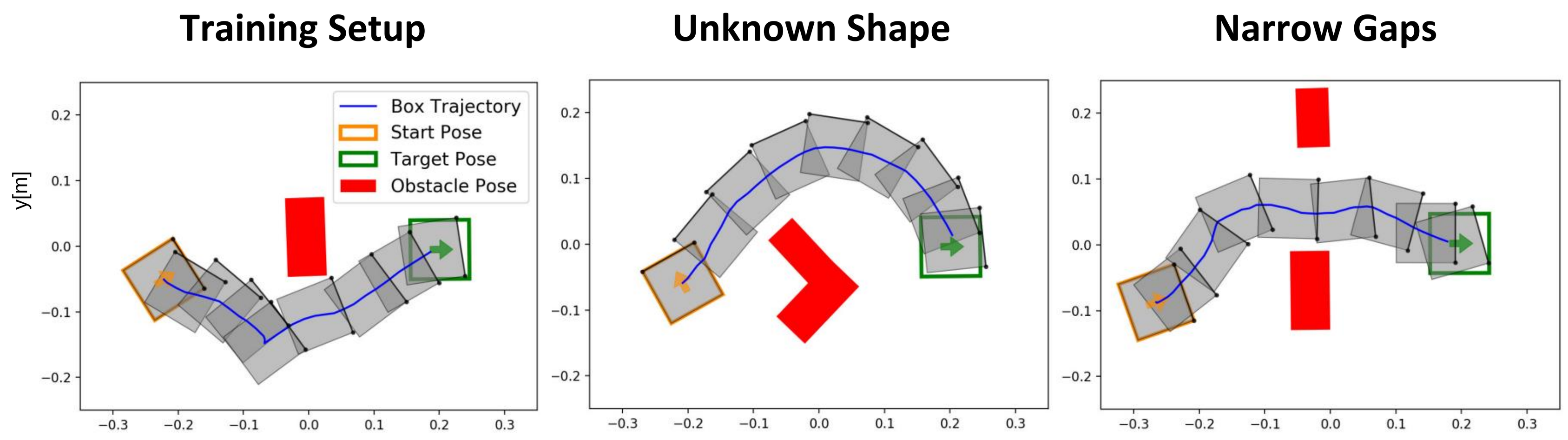
$$r_{total} = r_{term} + k_1(1 - r_{dist}) + k_2(1 - r_{ang}) + r_{coll}$$

Experimental Evaluation

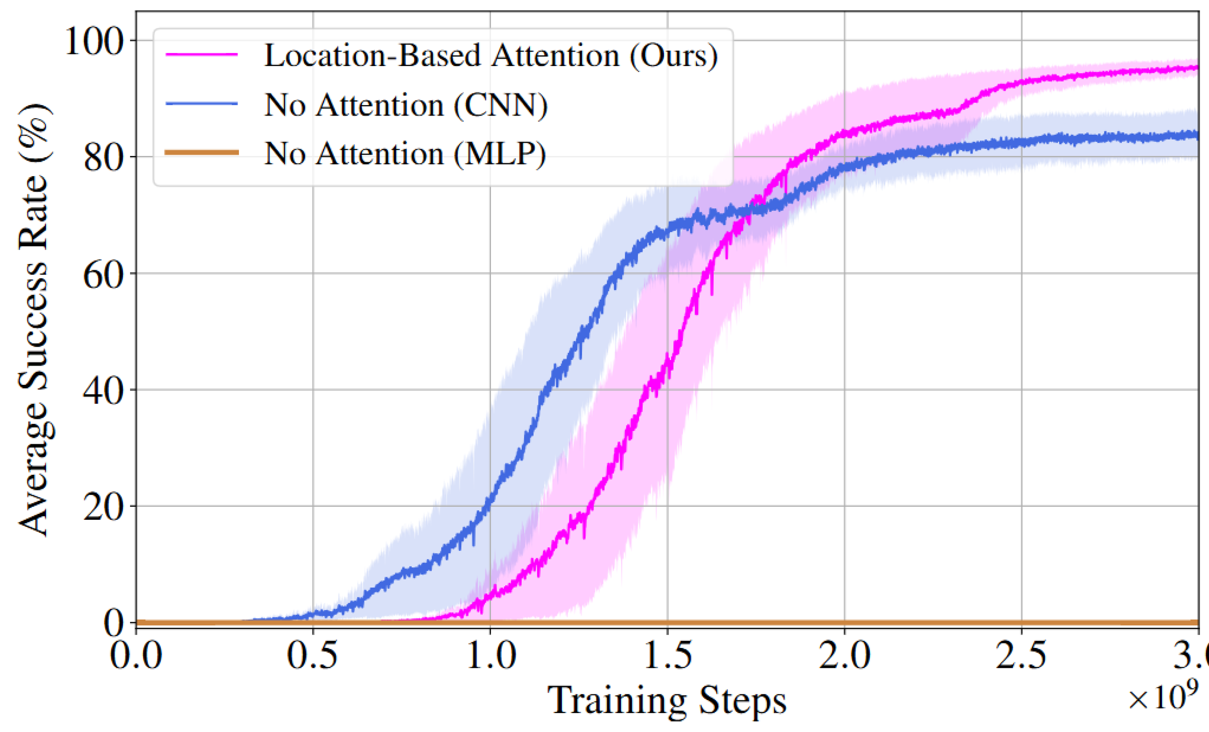
- Performance across different obstacle configurations in size, shape, and quantity conducted on 2,000 episodes
- High success rates across all configurations
- Slight increase in collision rates as complexity rises

Experimental Setup	Location Based Attention (Ours) Success Rate %	Location Based Attention (Ours) Collision Rate %	CNN Feature Extraction Success Rate %	CNN Feature Extraction Collision Rate %
Training	97.1	1.26	88.5	4.83
Circular	95.6	2.66	84.7	0.56
Cross-Shape	94.1	2.90	84.5	1.75
T-Shape	93.5	4.72	85.3	0.97
L-Shape	90.2	7.75	83.8	2.47
Dual Obstacles	48.1	50.7	57.9	34.3
Dual fine-tuned (DFT)	91.2	3.54	61.1	3.22
Circular (DFT)	96.4	0.20	72.1	0.34
Cross-Shape (DFT)	96.7	0.33	73.8	0.54
T-Shape (DFT)	96.3	1.32	71.9	1.01
L-Shape (DFT)	94.9	1.58	71.2	1.22

Qualitative Evaluation



Model Training



Real-World Validation

- Generalization to everday objects → not included in training

Summary

- Non-prehensile planar pushing in cluttered environments, leveraging location-based attention for improved feature extraction
- Validation in real-world scenario, demonstrating smooth and precise object pushing in challenging clutter layouts, including dynamic obstacles

Contact

hrl.uni-bonn.de

Nils Dengler
dengler@cs.uni-bonn.de
Humanoid Robots Lab
University of Bonn
Germany

Paper