Hindsight Experience Replay with Workspace Relabeling

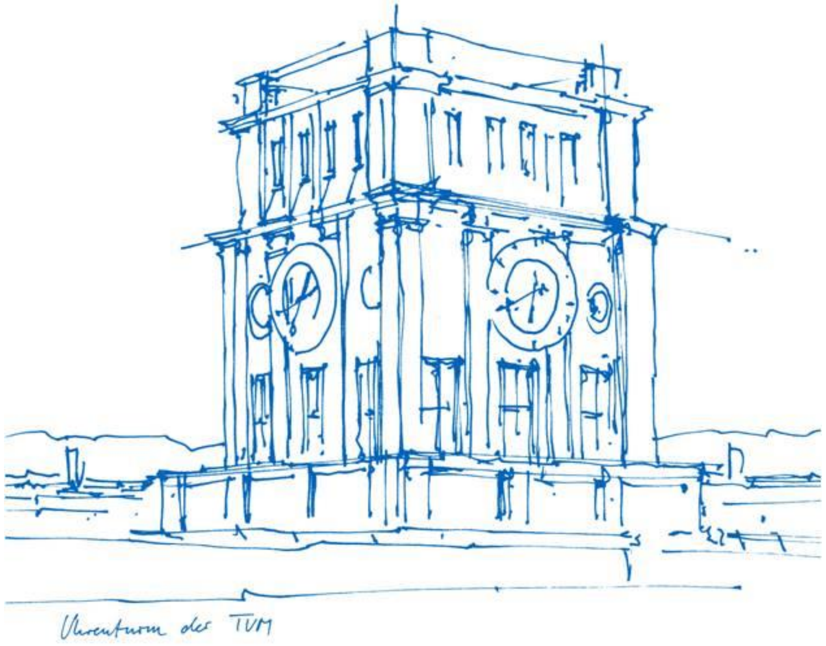
Supervisor:

Johannes Tenhumberg

Presenter:

Ferenc Török
Katharina Hermann

Advanced Deep Learning in Robotics (IN2349)
Technical University of Munich
Munich, 20.08.2020



Agenda

Motivation & Problem Statement

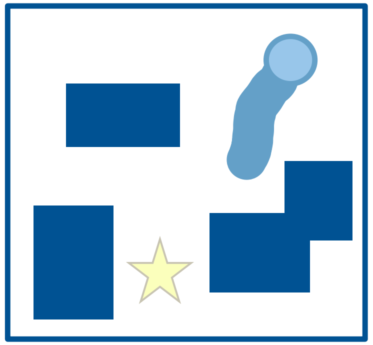
Goal:
Train a RL agent being able to find a trajectory from start to goal
through challenging and changing environments (workspace with obstacles)

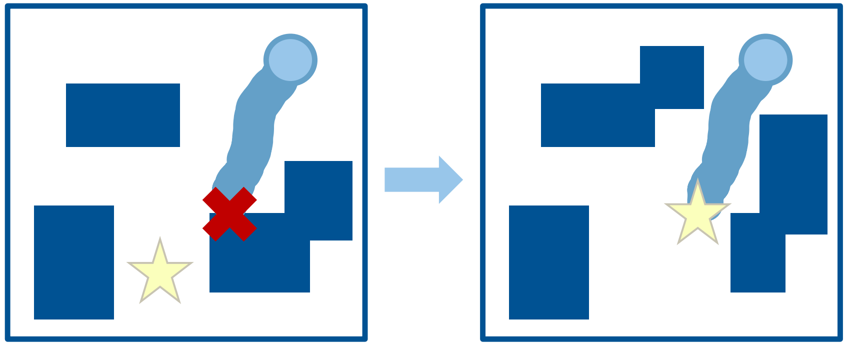
Idea:

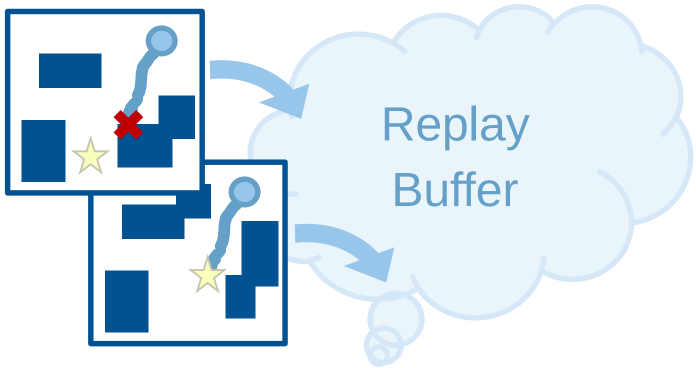
For each episode:
agent explores a new workspace
creating a movement trajectory

Trajectory fails:
Relabel as a success in hindsight
by modifiying the goal and the workspace

Save all trajectories (success and failures) as memories (replay buffer), which the robot agent can learn from







Related work

HER

Relabeling a failure trajectory (goal and reward) as a success in hindsight

Our approach:

Expert demonstrations via workspace relabeling

DDPG-MP with expert trajectories

No relabeling but inserting expert demonstrations via traditional MP algorithms in Replay buffer

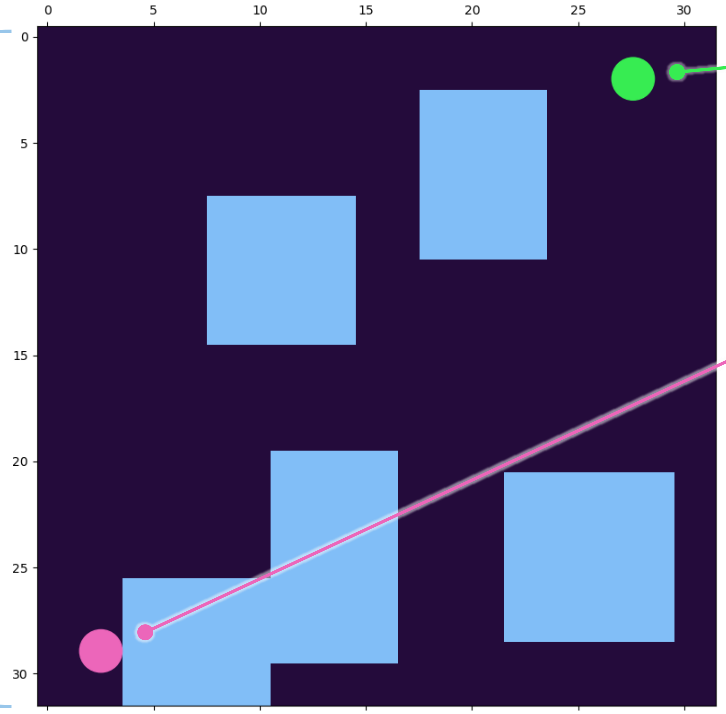
Agenda

1. Recap Motivation & Problem Statement
2. Relabeling Methods
3. Training and Test Setup
4. Results
5. Summary & Outlook

Experiment setup

Workspace:
- 32x32 grid with random obstacles

Reduced workspace:
- WS representation reduced through a Convolutional Autoencoder to $\{\mathrm{w}\}$.



Start (pink) & Goal (yellow)
- Created randomly for a given WS

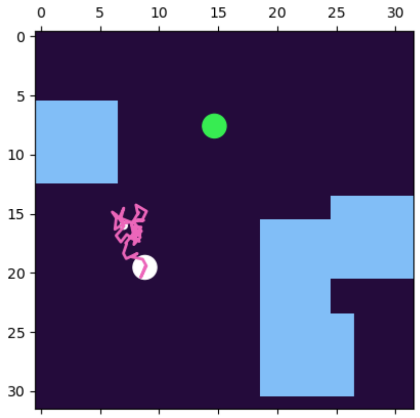
Agent

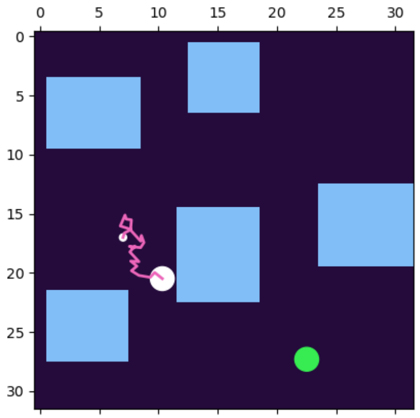
2D circle robot of radius 1 with:
- a_t: 2 DOF movement in x and y direction
- $s_t = \{\mathrm{c_t}, \mathrm{w}, \mathrm{g}\}$
- $r_t^T = \begin{cases} -0{,}06 & (c_t, c_{t+1}) \in T_{free}, \\ 8 & (c_t, c_{t+1}) \in T_{goal}, \\ -7 & (c_t, c_{t+1}) \in T_{col}. \end{cases}$

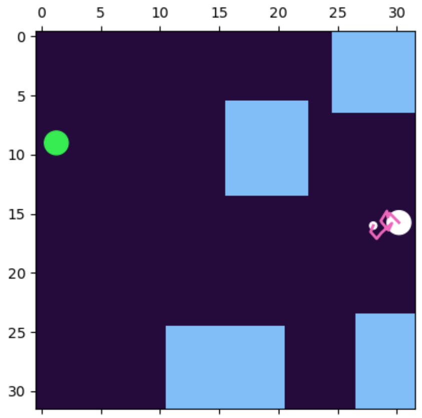
Relabeling methods: Cutting trajectory - Warmup

Max steps | **Obstacle collision** | **Wall collision**

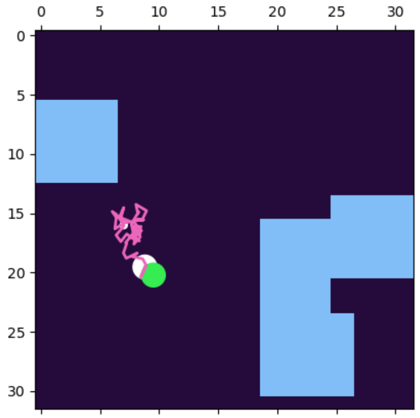
Before

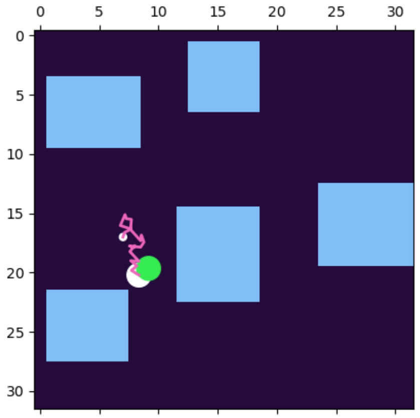


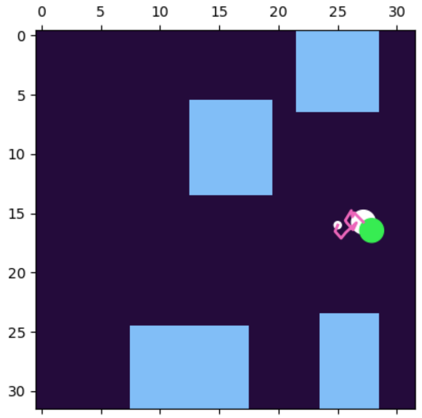




After

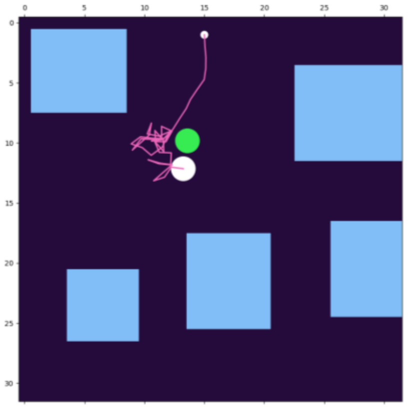
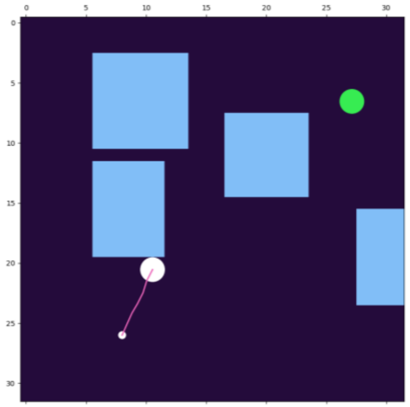
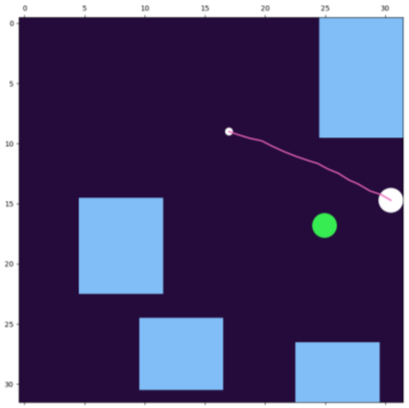






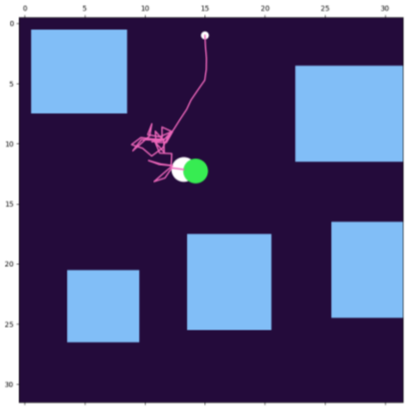
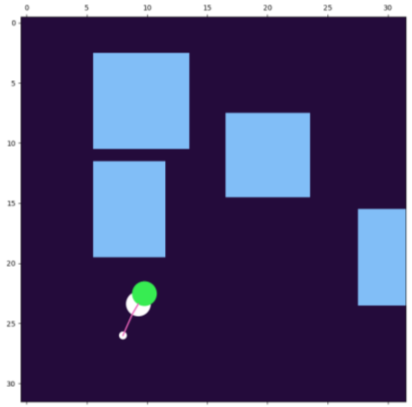
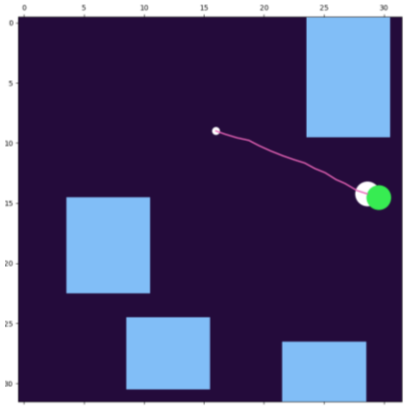
Relabeling methods: Cutting trajectory - Training

Max steps | **Obstacle collision** | **Wall collision**

Before

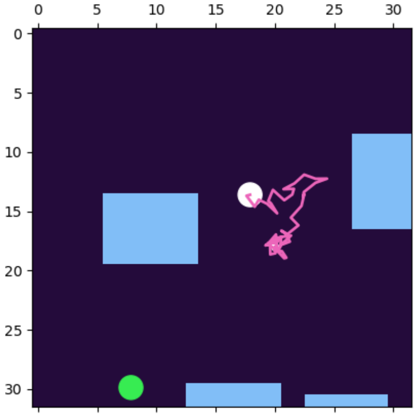
After

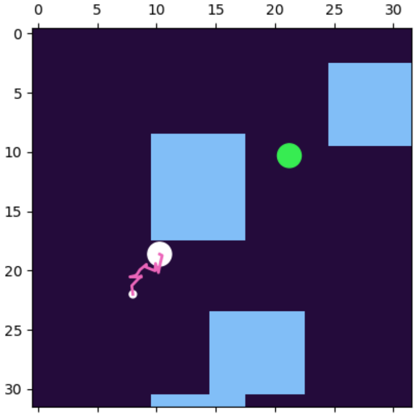




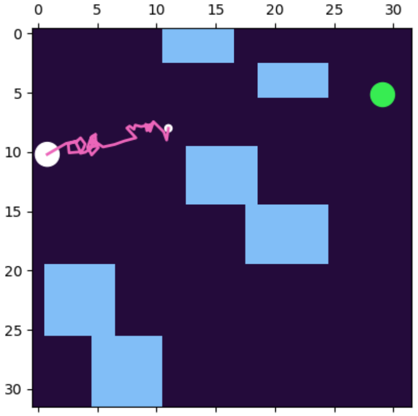
Relabeling methods: Random passage - Warmup

Max steps | **Obstacle collision** | **Wall collision**

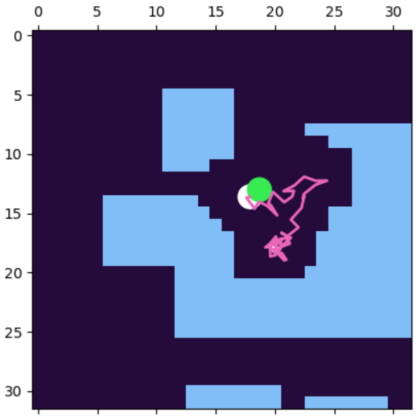
Before

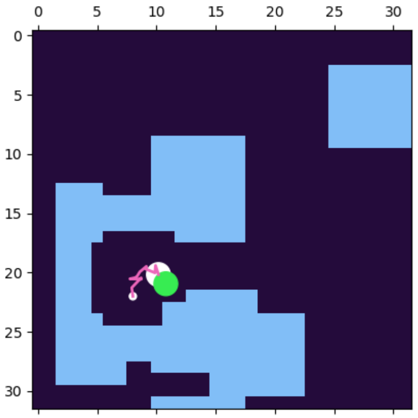


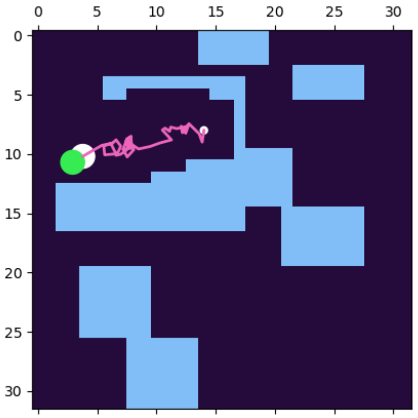




After





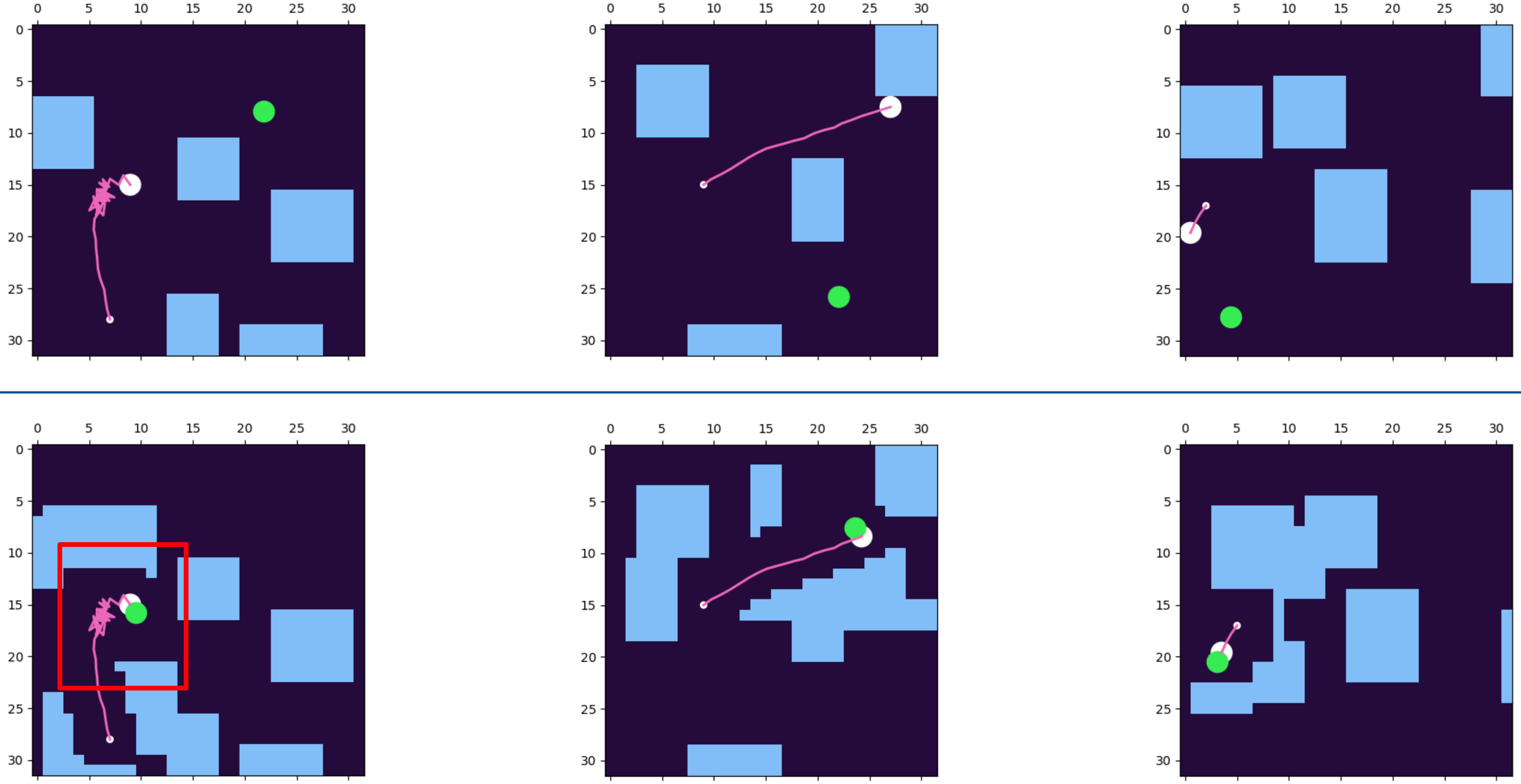


Relabeling methods: Random passage - Training

Max steps | **Obstacle collision** | **Wall collision**

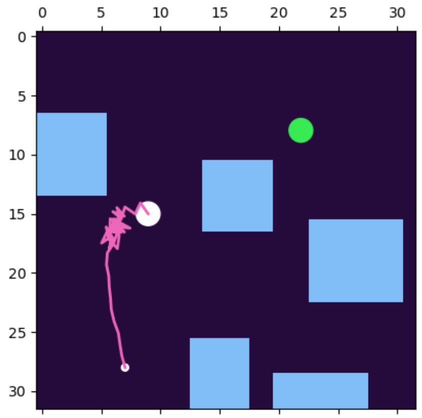
Before

After



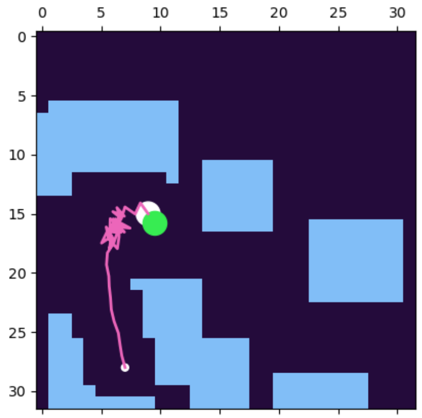
Relabeling methods – Jitter movement

Zig-Zag trajectories



Before

After



Problem

- Suboptimal trajectories relabeled and saved in RB as „perfect“ samples
- Robot will not learn to avoid jitter movement

Idea

- Check Zig-Zaging of a trajectory (avg angle)
- Train with & without removement of Zig-Zag-trajectories from relabeling
- Compare results

Agenda

Baseline training in WS without obstacles

Training parameters

Workspace:

- **Empty WS**
- No maximun goal distance

Actor and Critic network architecture:

- 2 layer (400 & 300 hidden units)

Training:

- 800 k steps (max 50 steps/ episode)
- Warmup: 10k steps
- Policy Update every step
- Evaluation every 10k steps (for 100 episodes)
- Weight decay of 0.001
- Batch size: 256

Relabeling method	Zig-Zag removement	
No-relabeling		1
Straight line relabeling		2
Cutting trajectory	Yes	3
	No	4

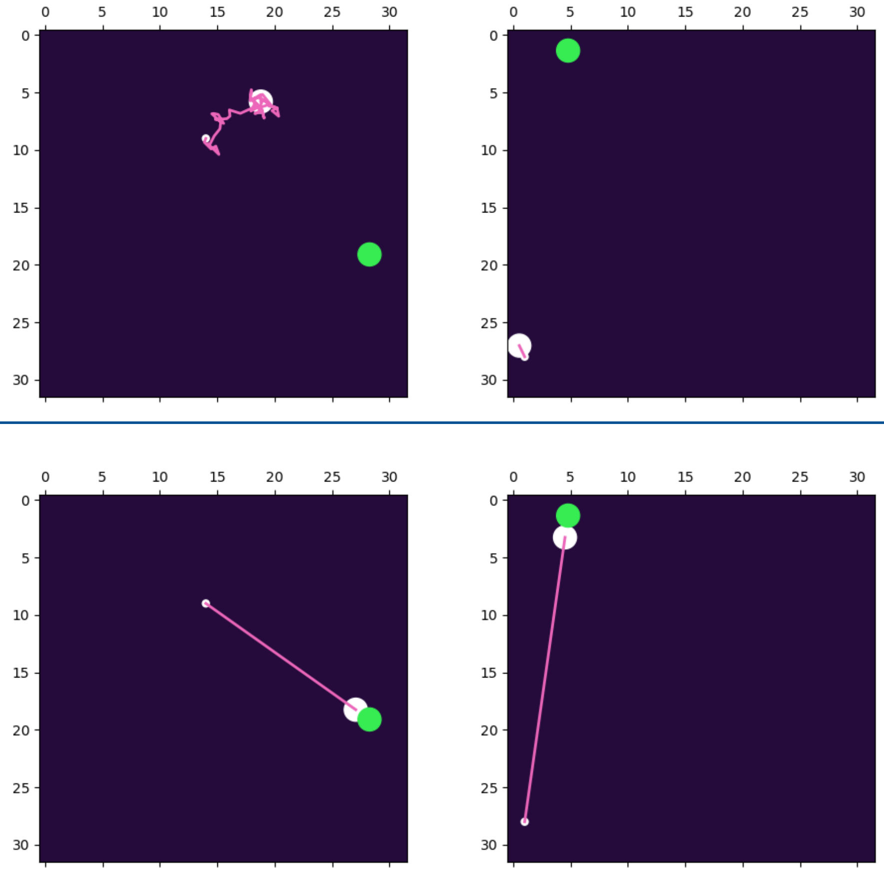
Baseline methods: Expert trajectories vs Relabeling

Straight line

Max steps / Wall collision

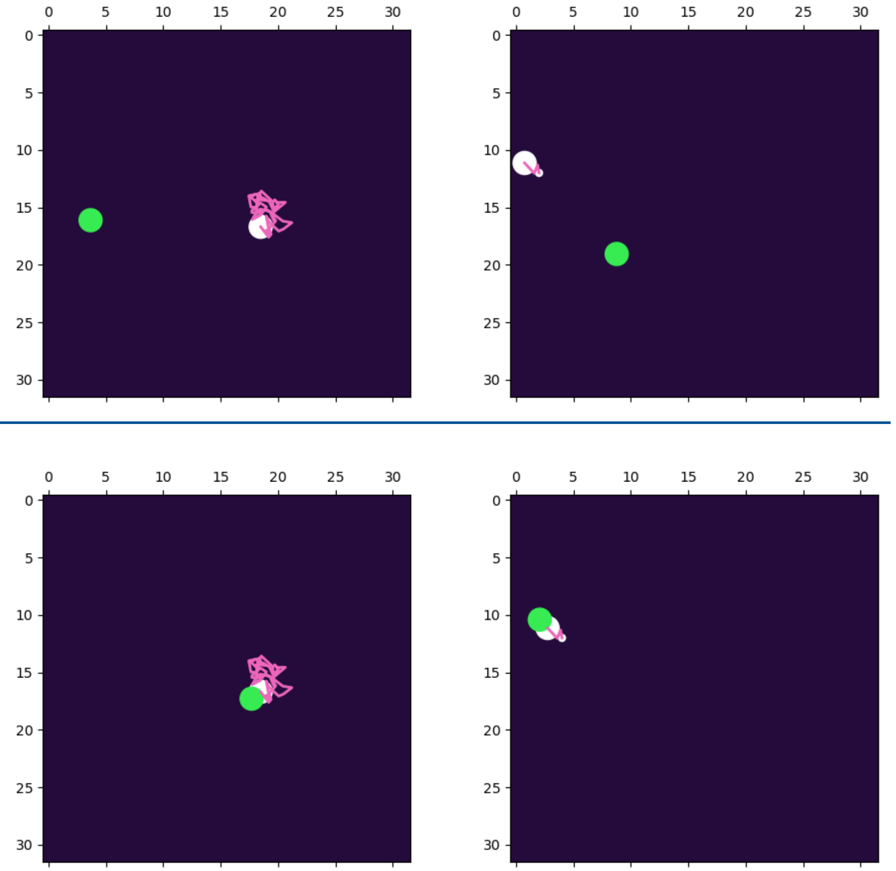
Before

After



Cutting trajectory

Max steps / Wall collision



Training in WS with obstacles

Training parameters

Workspace:

- **Mid level: 4 - 5 obstacles & avg_size 8** (buffer with 100 WS)
- No maximun goal distance

Actor and Critic network architecture:

- 2 layer (400 & 300 hidden units)

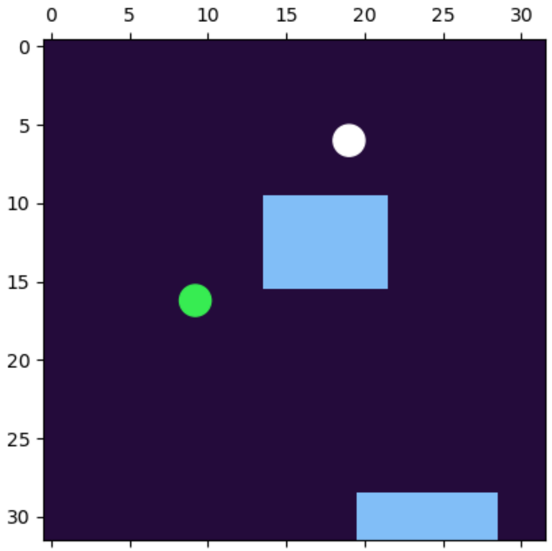
Training:

- 800 k steps (max 50 steps/ episode)
- Warmup: 10k steps
- Policy Update every step
- Evaluation every 10k steps (for 100 episodes)
- Weight decay of 0.001
- Batch size: 256

Relabeling method	Zig-Zag removement	
No-relabeling		1
Cutting trajectory	Yes	2
Cutting trajectory	No	3
Random passage	Yes	4
Random passage	No	5

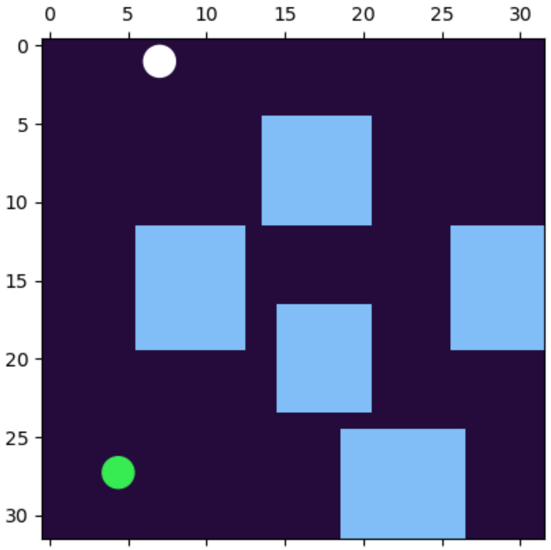
Evaluation on 3 different workspace levels

Easy



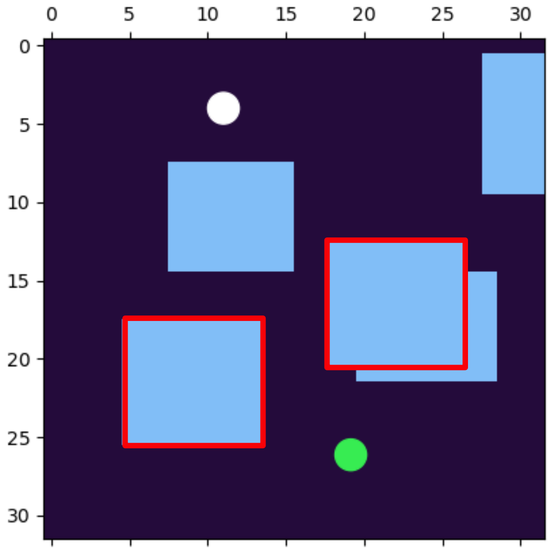
- 2-3 objects
- Avg objects size: 8 (normal distributed)
- Object centers uniformly distibuted
- Max goal distance: 15

Medium



- 4 - 5 objects
- Avg objects size: 8 (normal distributed)
- Object centers uniformly distibuted

Hard

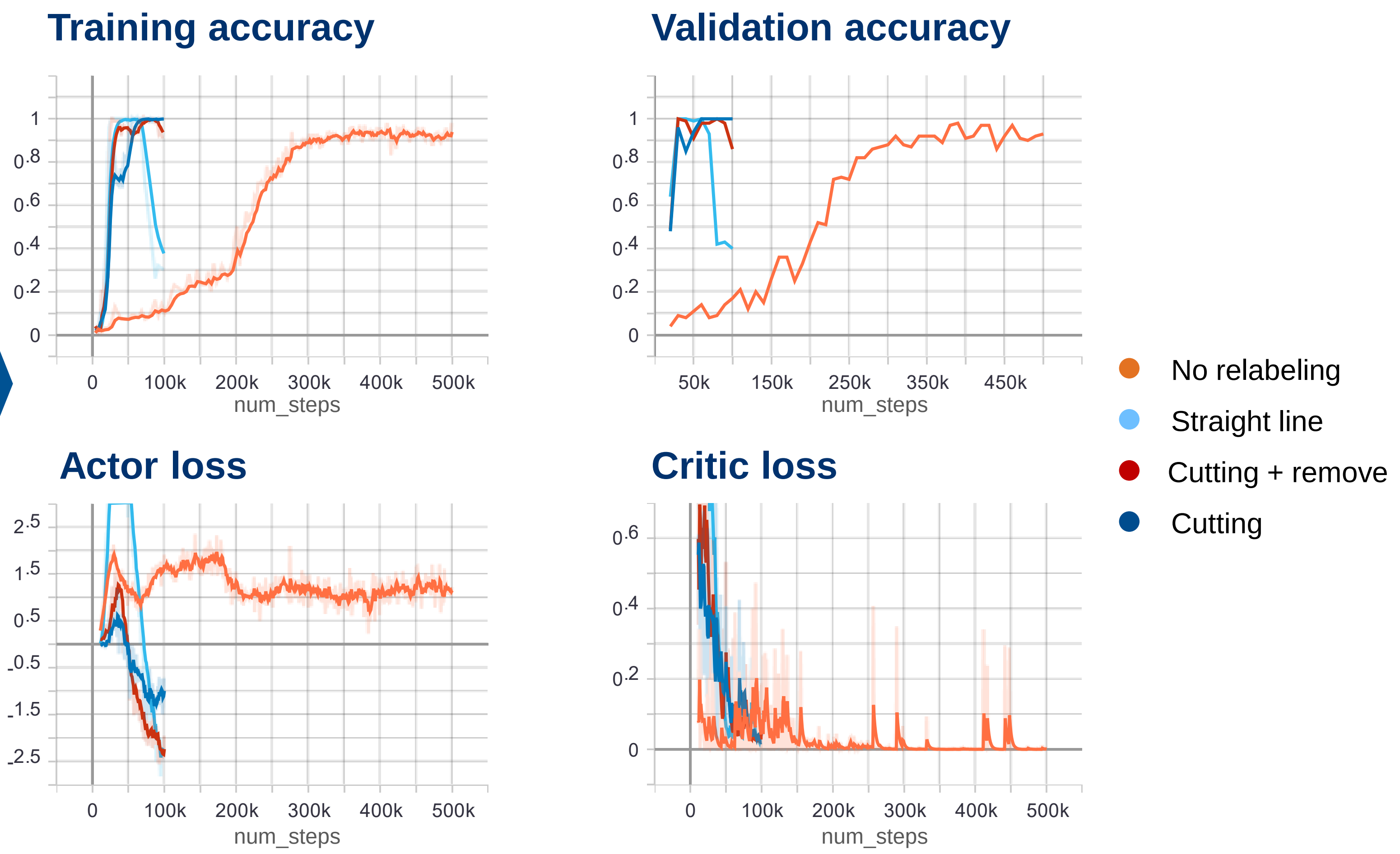


- 5 - 6 objects
- Avg objects size: 8
- Object centers uniformly distibuted
- Narrow passage of 2 objects

Agenda

Baseline - Empty WS Training and Evaluation

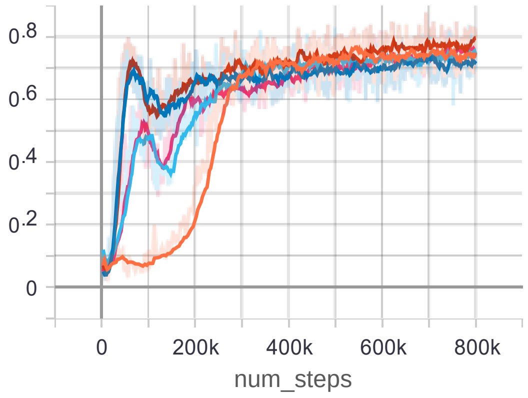
Relabeling method	Zig-Zag removement	
No relabeling		1
Straight line relabeling		2
Cutting trajectory	Yes	3
	No	4



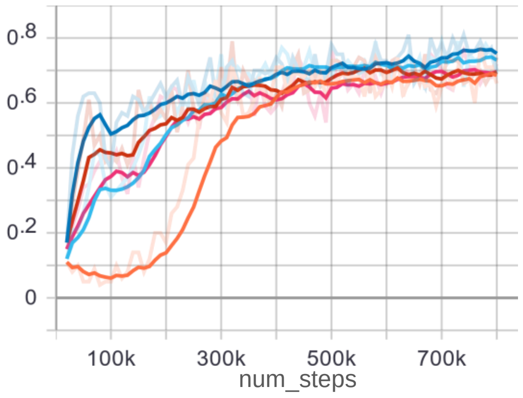
Obstacle WS Training

Relabeling method	Zig-Zag removement	
No relabeling		1
Cutting trajectory	Yes	2
	No	3
Random passage	Yes	4
	No	5

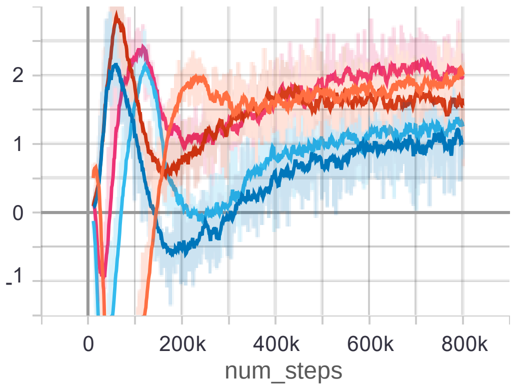
Training accuracy



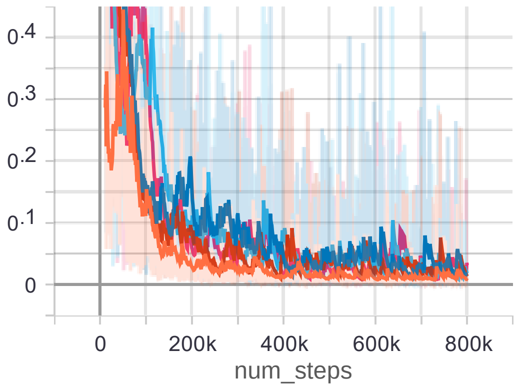
Validation accuracy



Actor loss



Critic loss



- No relabeling
- Cutting + remove
- Cutting
- Random + remove
- Random

Obstacle WS Evaluation: Easy

Relabeling method	Zig-Zag removement	#	Straight line solution feasibility	Test accuracy: Total	Test accuracy: Straight lines	Test accuracy: No straight lines
No relabeling		1	89% of 1000 episodes	88%	91%	62%
Cutting trajectory	Yes	2	90% of 1000 episodes	96%	99%	70%
Cutting trajectory	No	3	87% of 1000 episodes	95%	99%	65%
Random passage	Yes	4	89% of 1000 episodes	94%	98%	61%
Random passage	No	5	87% of 1000 episodes	90%	95%	60%

Obstacle WS Evaluation: Medium

Relabeling method	Zig-Zag removement		Straight line solution feasibility	Test accuracy		
				Total	Straight lines	No straight lines
No relabeling		1	55% of 1000 episodes	62%	91%	25%
Cutting trajectory	Yes	2	58% of 1000 episodes	64%	93%	23%
Cutting trajectory	No	3	54% of 1000 episodes	63%	94%	27%
Random passage	Yes	4	59% of 1000 episodes	70%	97%	31%
Random passage	No	5	59% of 1000 episodes	67%	93%	30%

Obstacle WS Evaluation: Hard

Relabeling method	Zig-Zag removement		Straight line solution feasibility	Test accuracy		
				Total	Straight lines	No straight lines
No relabeling		1	1% of 1000 episodes	31%	100%	30%
Cutting trajectory	Yes	2	1% of 1000 episodes	32%	75%	32%
Cutting trajectory	No	3	1% of 1000 episodes	34%	100%	34%
Random passage	Yes	4	1% of 1000 episodes	40%	100%	40%
Random passage	No	5	1% of 1000 episodes	38%	80%	38%

Agenda

Summary

- Slight improvement with relabeling over no relabeling: comparable to HER
- Relabeling success limited due to zig-zaging or too easy (straight-line) trajectories.

Comparison to DDPG-MP with expert trajectories:

- Not as good results since expert trajectories are more optimal to solve complex tasks

Outlook

- Creating more complex trajectories (but no jitter movement) first and a challenging environment around it

Thanks for your attention!