Comparison of Reinforcement Learning for Direct and Indirect Locomotion Control in Target Tracking with Snake-like Robots

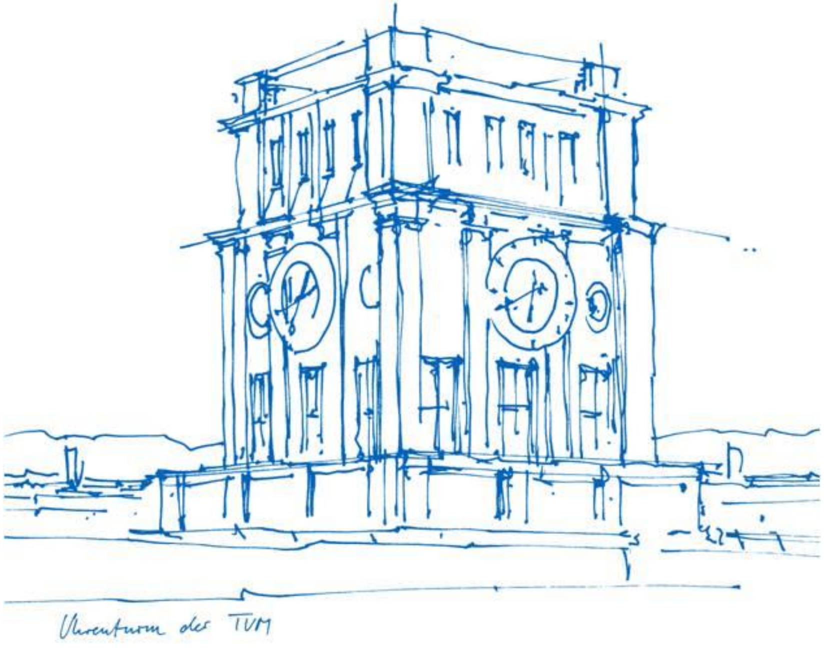
Julian Schmitz

Technical University of Munich

Department of Informatics

Bachelor‘s Thesis

Munich, 05. October 2018

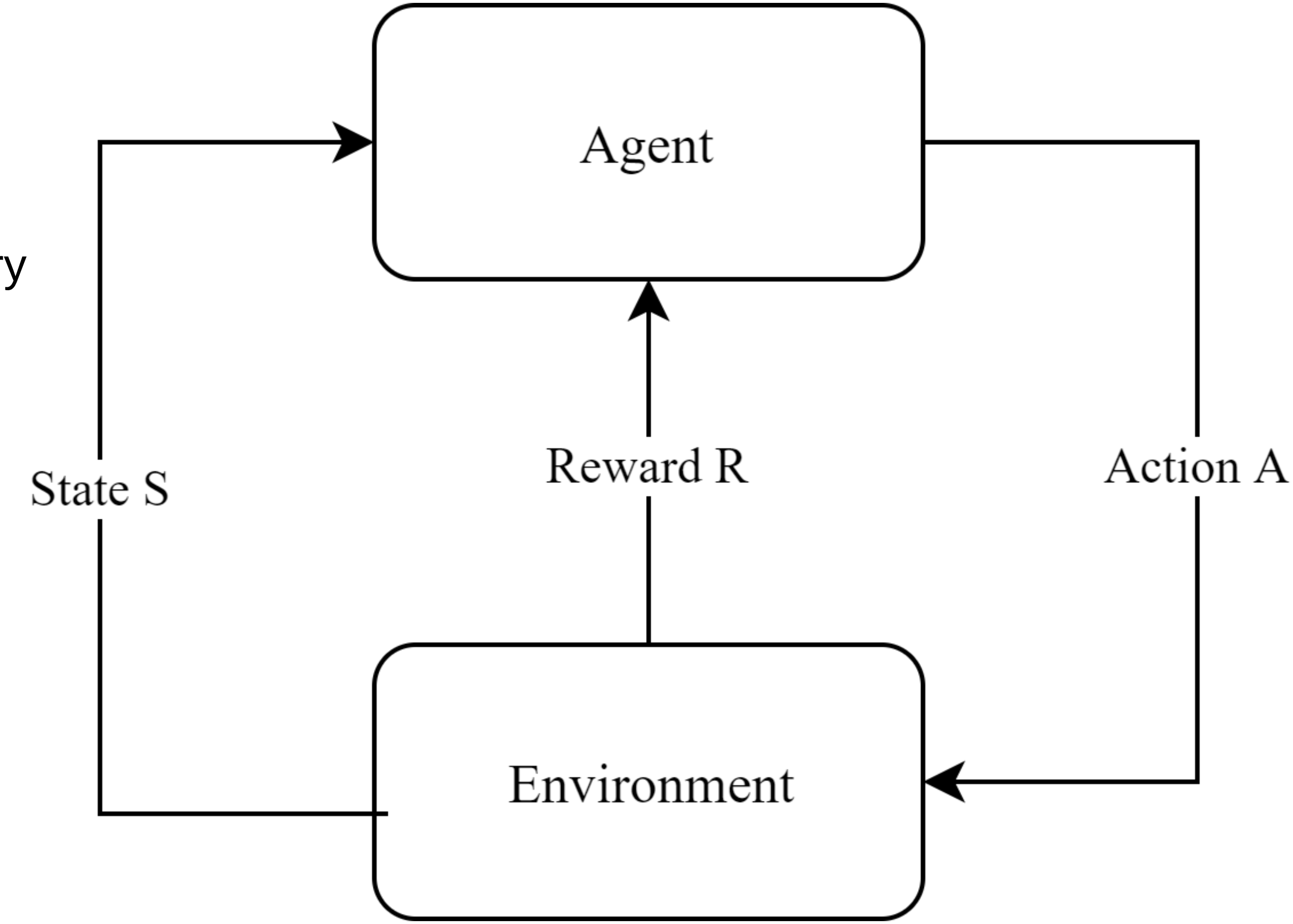


Agenda

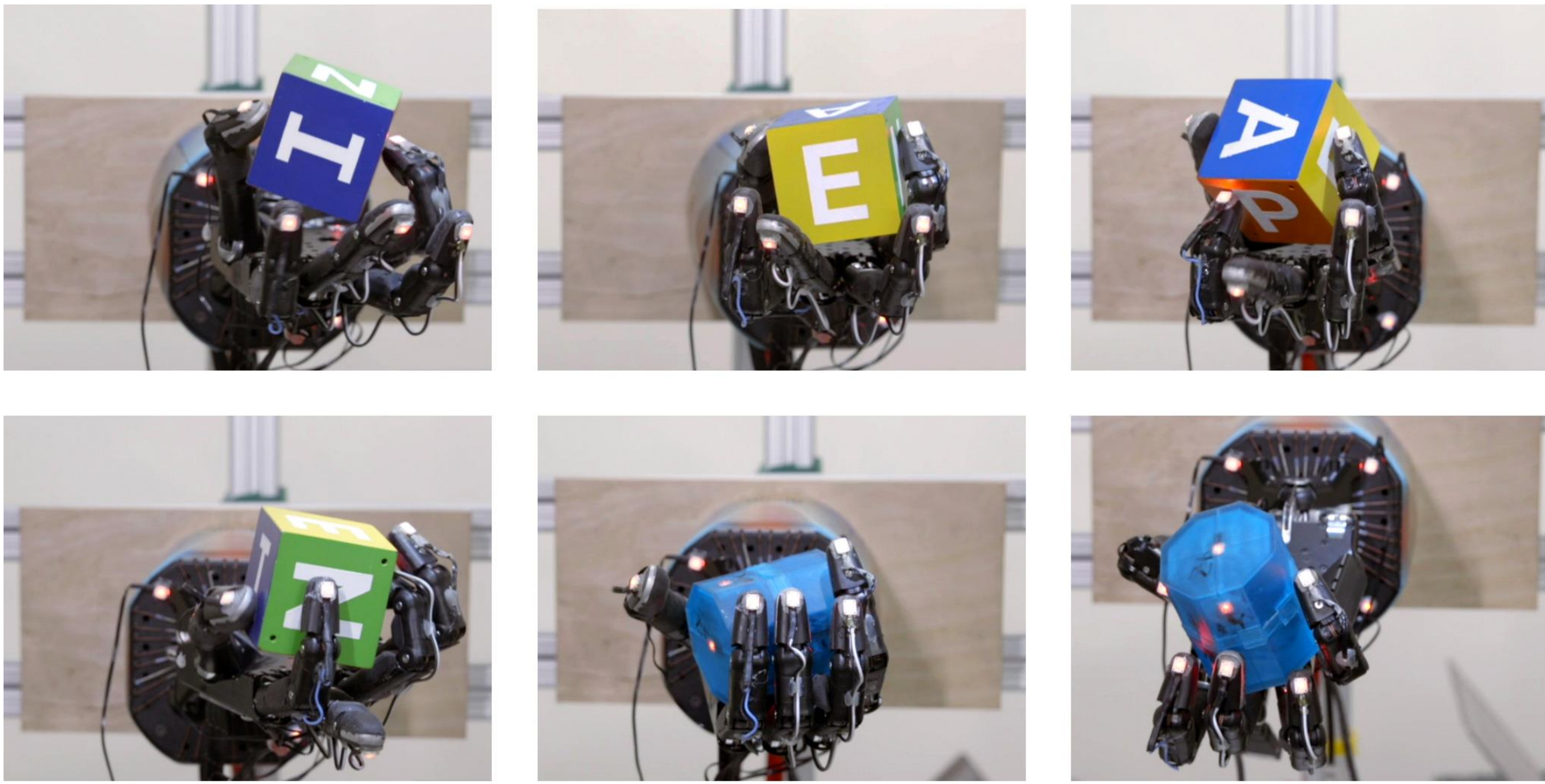
- Motivation
- Methodology
- Approach
- Results
- Conclusion

Motivation – Reinforcement Learning

- Define reward
- Learn by discovery
- Assess goal



Motivation - Reinforcement Learning in Robotics

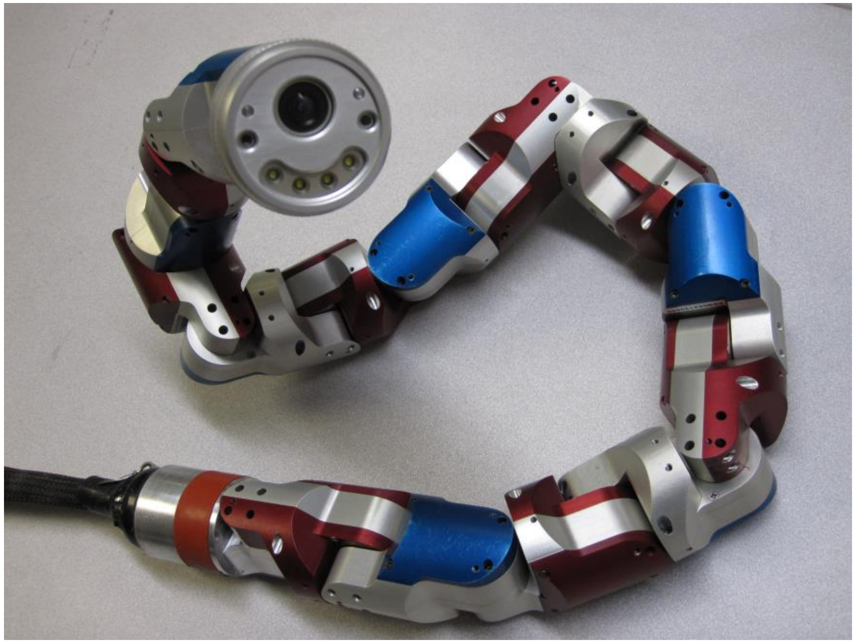


https://blog.openai.com/learning-dexterity/

Motivation - Reinforcement Learning in Robotics

https://wayve.ai/blog/learning-to-drive-in-a-day-with-reinforcement-learning

Motivation - Snake-like robots



http://biorobotics.ri.cmu.edu/projects/modsnake/pictures.html

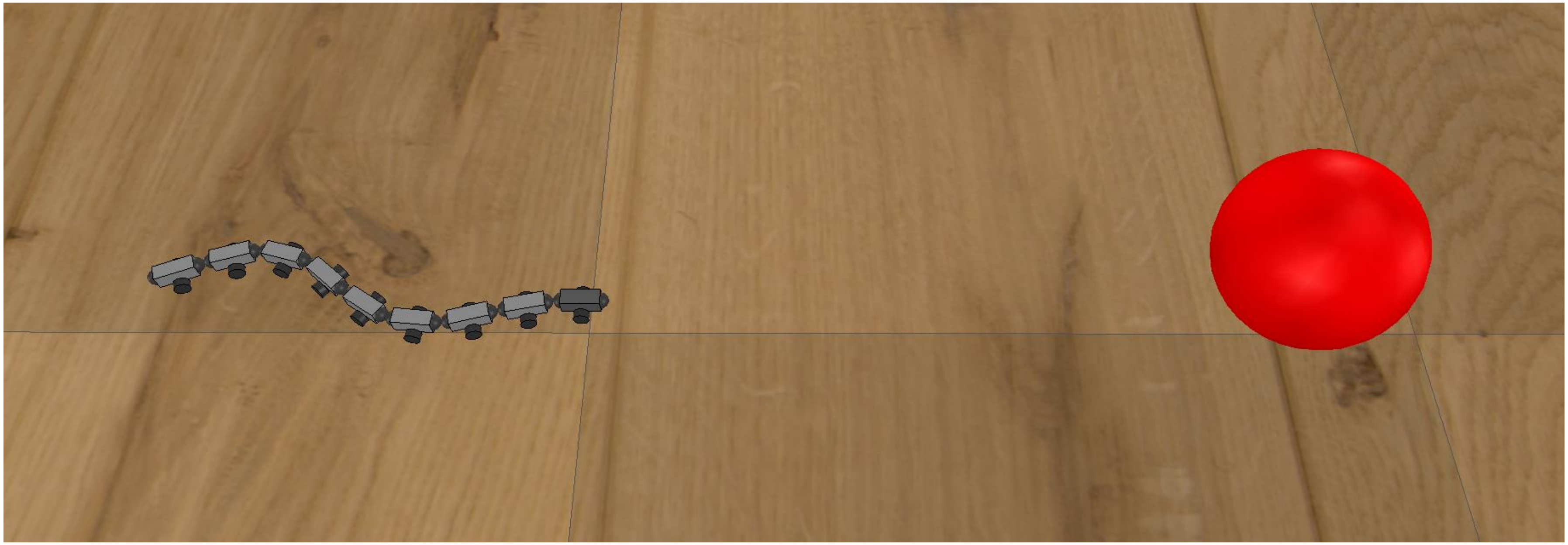
https://biorob.epfl.ch/salamandra

- Small diameter
- Good locomotion capabilities

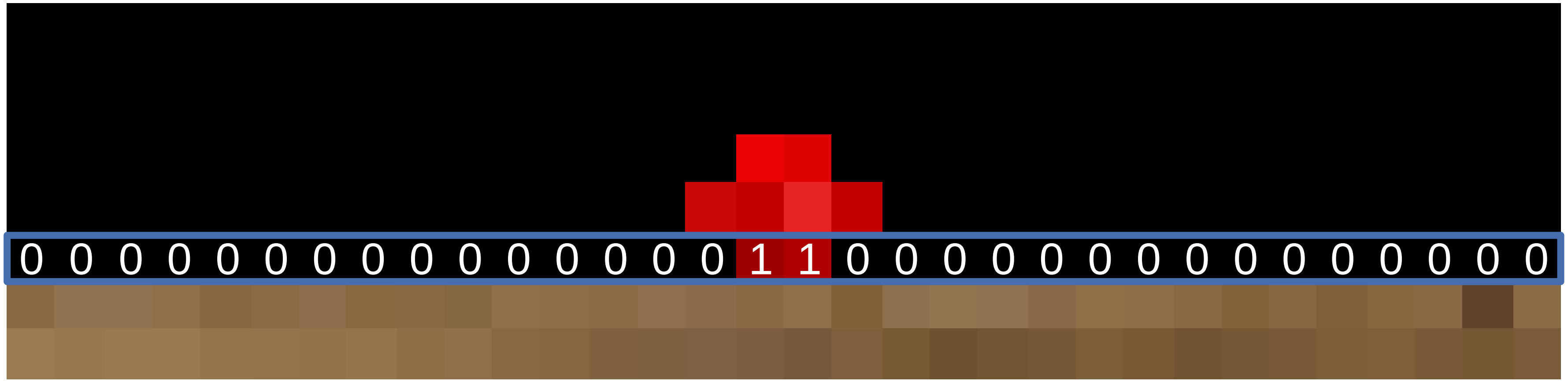
Agenda

- Motivation
- Methodology
- Approach
- Results
- Conclusion

Methodology - Scene



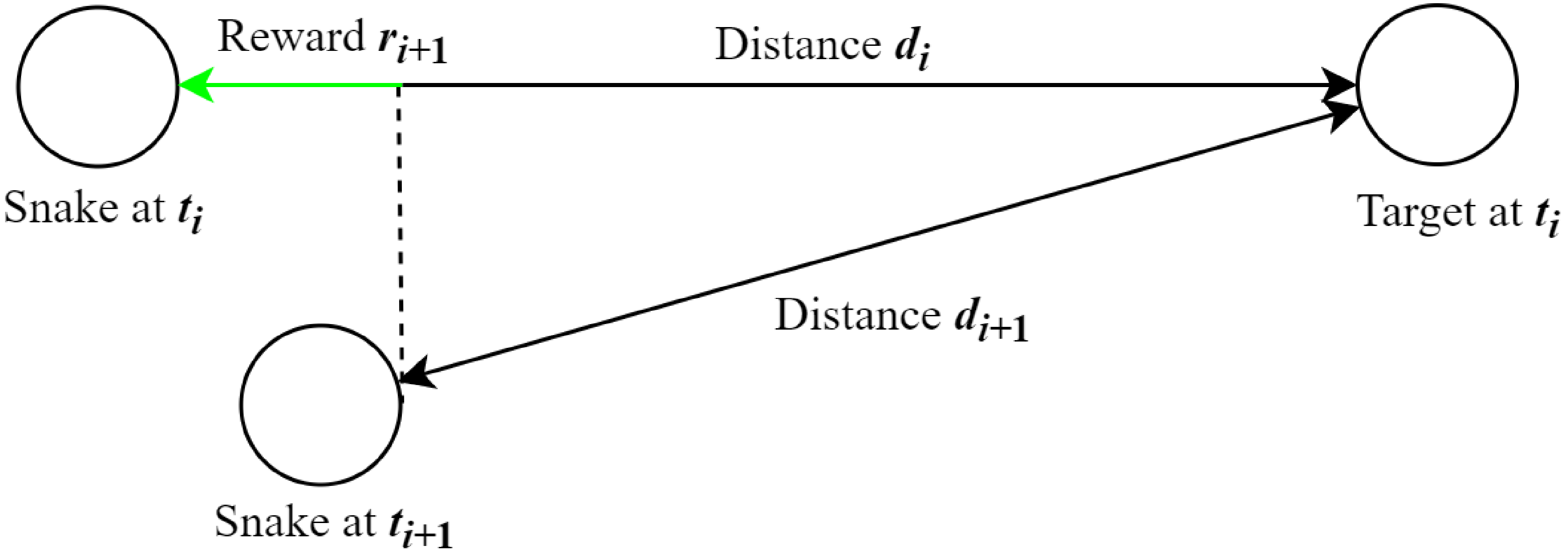
Methodology - Observation



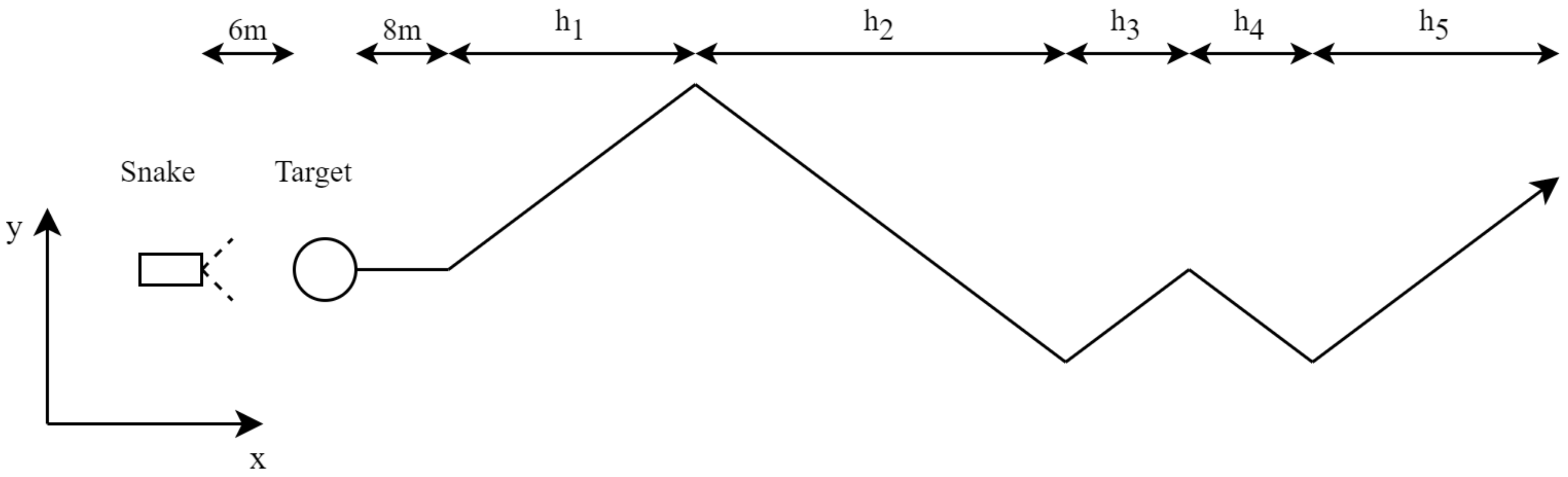
24 rows cropped from this image

Element	Observation Size
Vision sensor image	32
Current joint angles	8
Target joint angles	8
Head module speed	1
Total	**49**

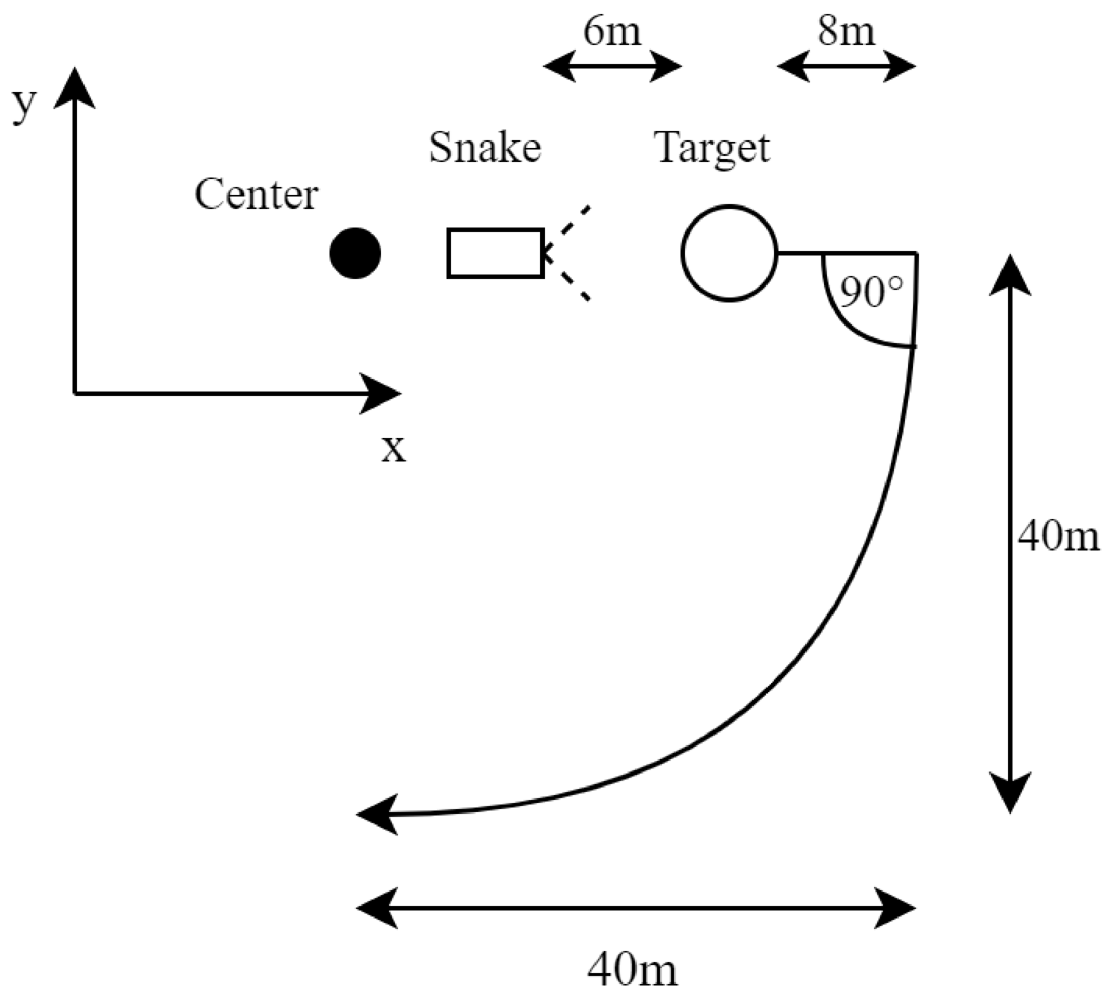
Methodology - Reward



Methodology – Training Scenario



Methodology – Evaluation Scenario



Methodology - Proximal Policy Optimization (PPO)

$$\max_{\theta} \hat{E}$$

Traditional Policy Gradient Loss

$$L^{PG}(\theta) = \hat{E}_t\left[\log \pi_\theta(a_t|s_t)\hat{A}_t\right]$$

Probabilities of output of the policy network

Estimate value of this output

Estimate > Average → Increase Probability

Problem: Destructively large policy updates

Trust Region Methods

$$L^{PG}(\theta) = \hat{E}_t\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}\hat{A}_t\right]$$

$r(\theta)$

PPO clips $r(\theta)$ between $1-\epsilon$ and $1+\epsilon$

- Easy implementation
- Relatively sample efficient
- Avoid high policy updates

Agenda

- Motivation
- Methodology
- Approach
- Results
- Conclusion

Approach – Indirect Locomotion Control

Observation

↓

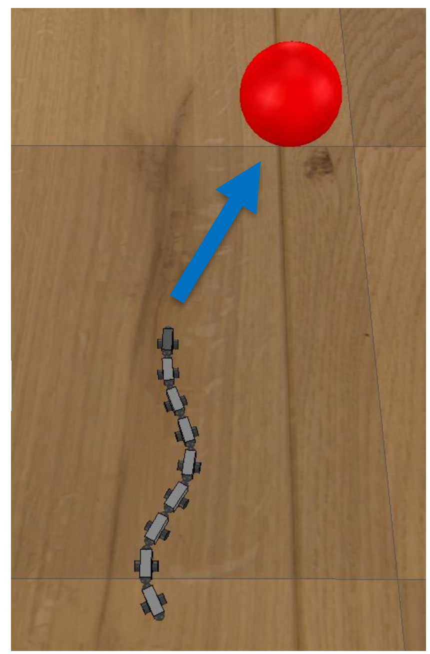
Reinforcement Learning Agent

↓

Direction

Approach – Indirect Locomotion Control

Proximal Policy Optimization



Locomotion by
Slithering gait [2]

[2] Shigeo Hirose. Biologically inspired robots : snake-like locomotors and manipulators.

Approach – Direct Locomotion Control

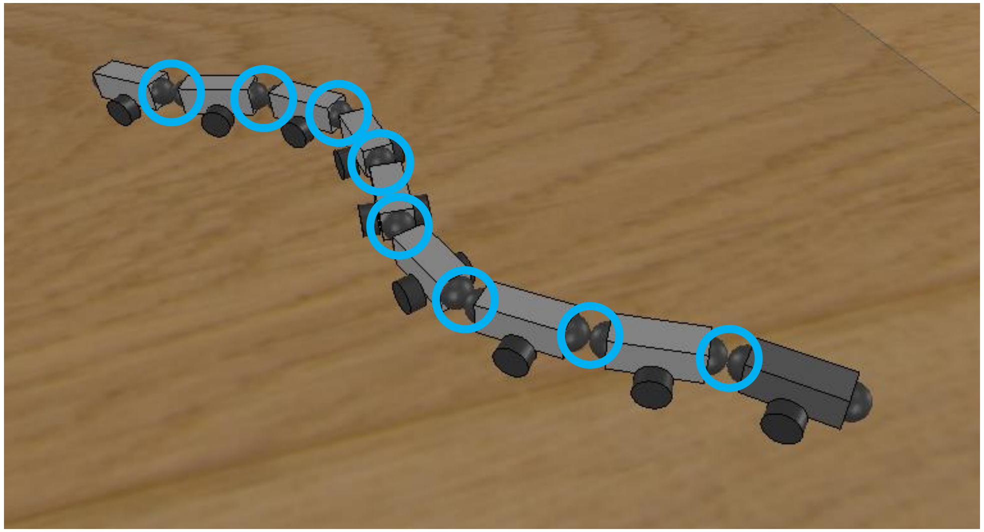
Observation

Reinforcement Learning Agent

Locomotion

Approach – Direct Locomotion Control

Proximal Policy Optimization

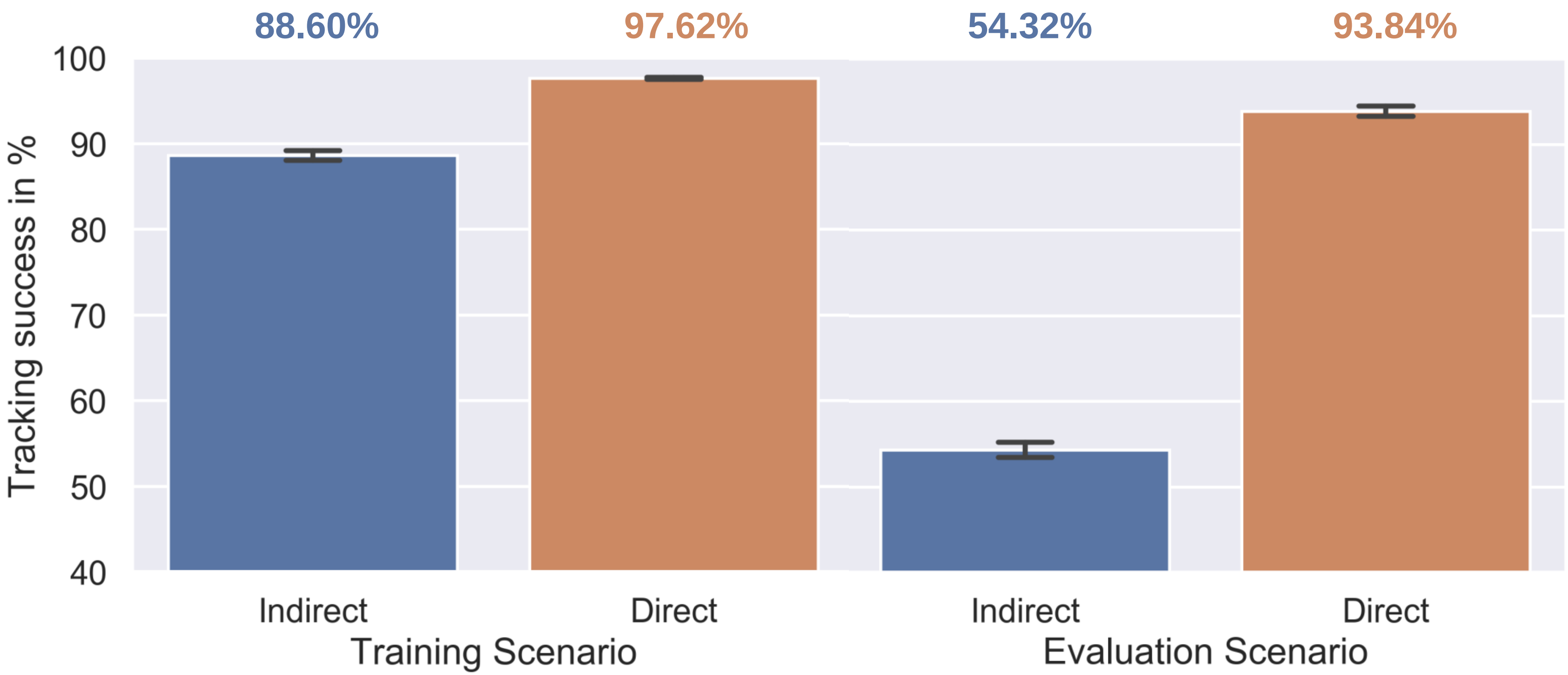


Action Space	8
Joint limits	-100°, 100°

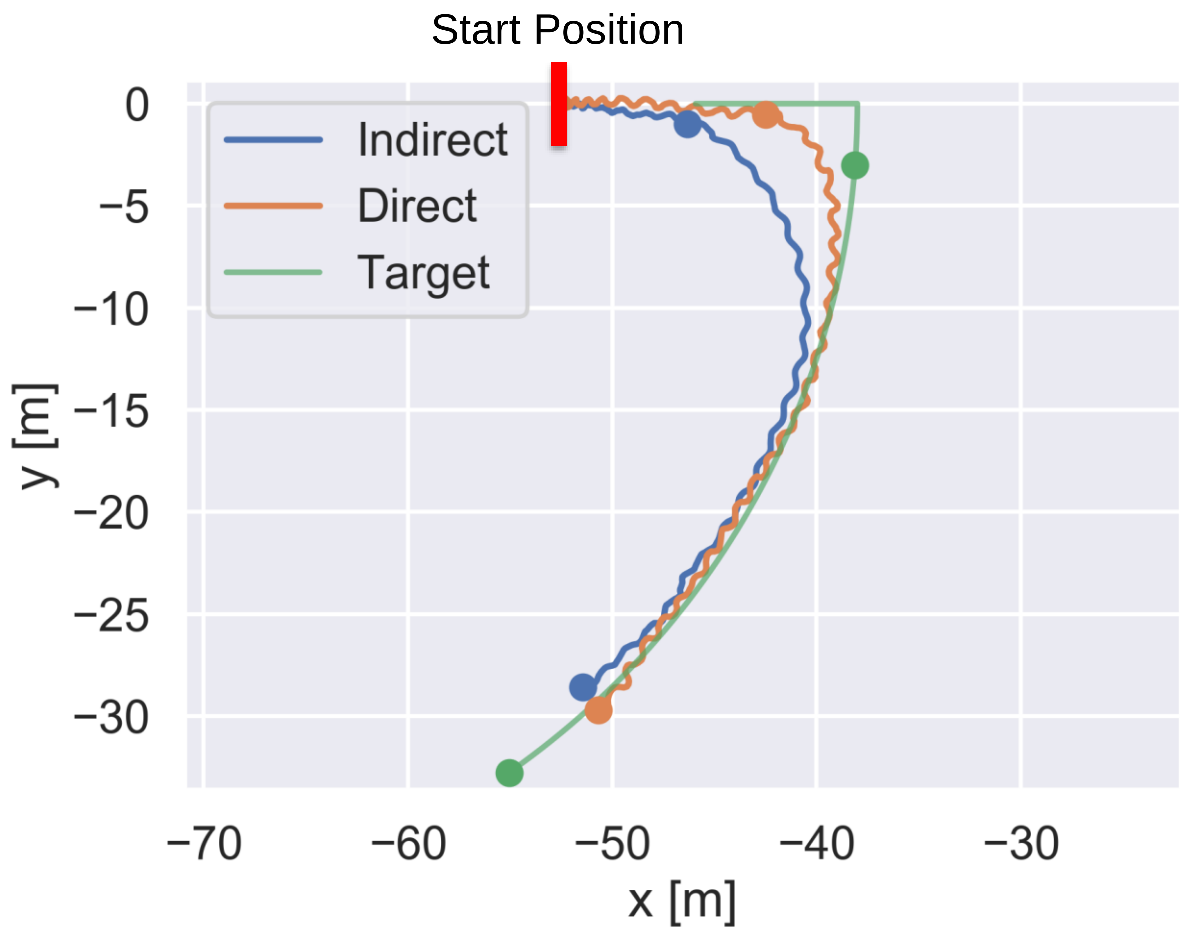
Agenda

- Motivation
- Methodology
- Approach
- Results
- Conclusion

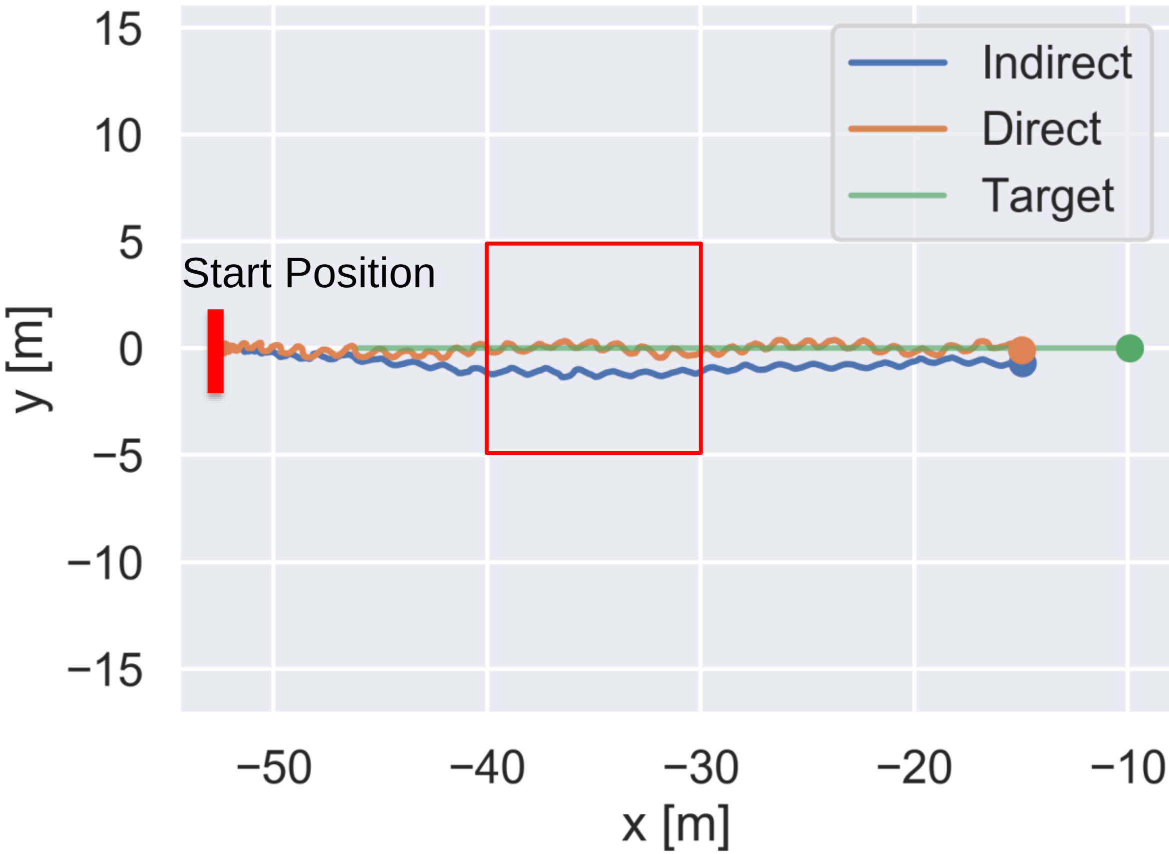
Results - Tracking Accuracy

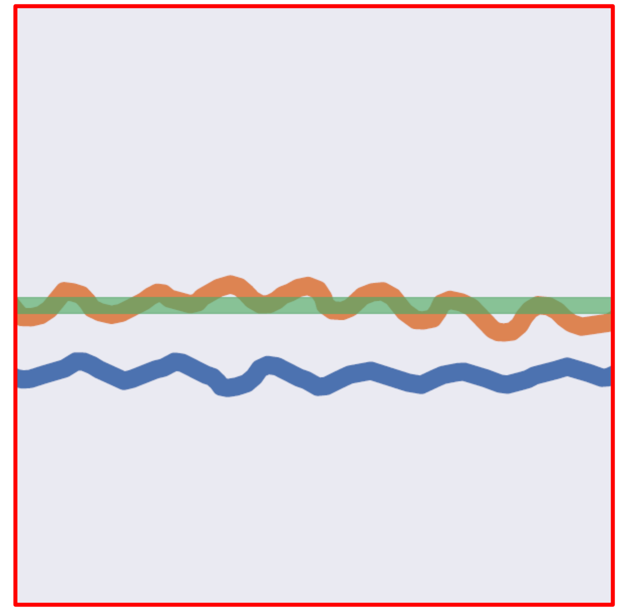


Results - Comparison

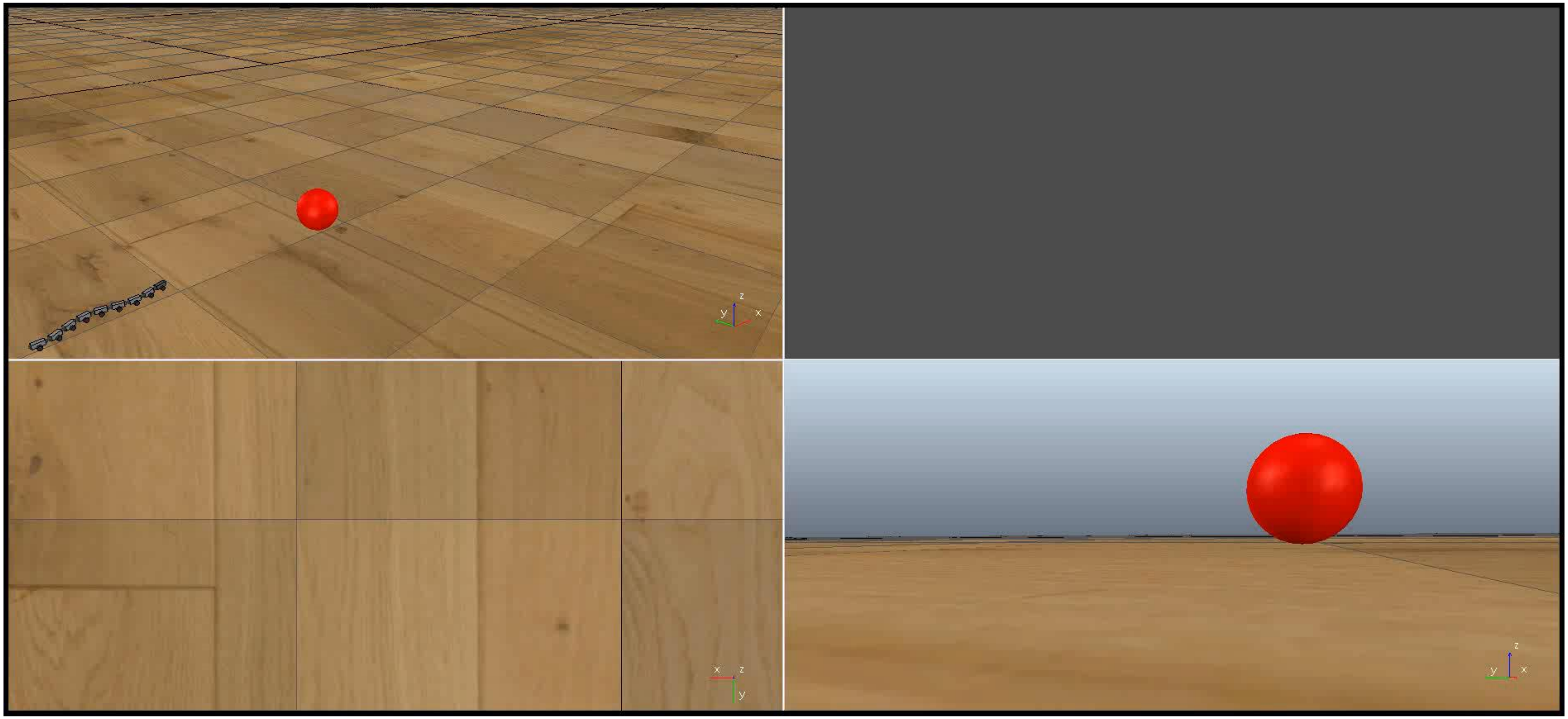


Results - Comparison

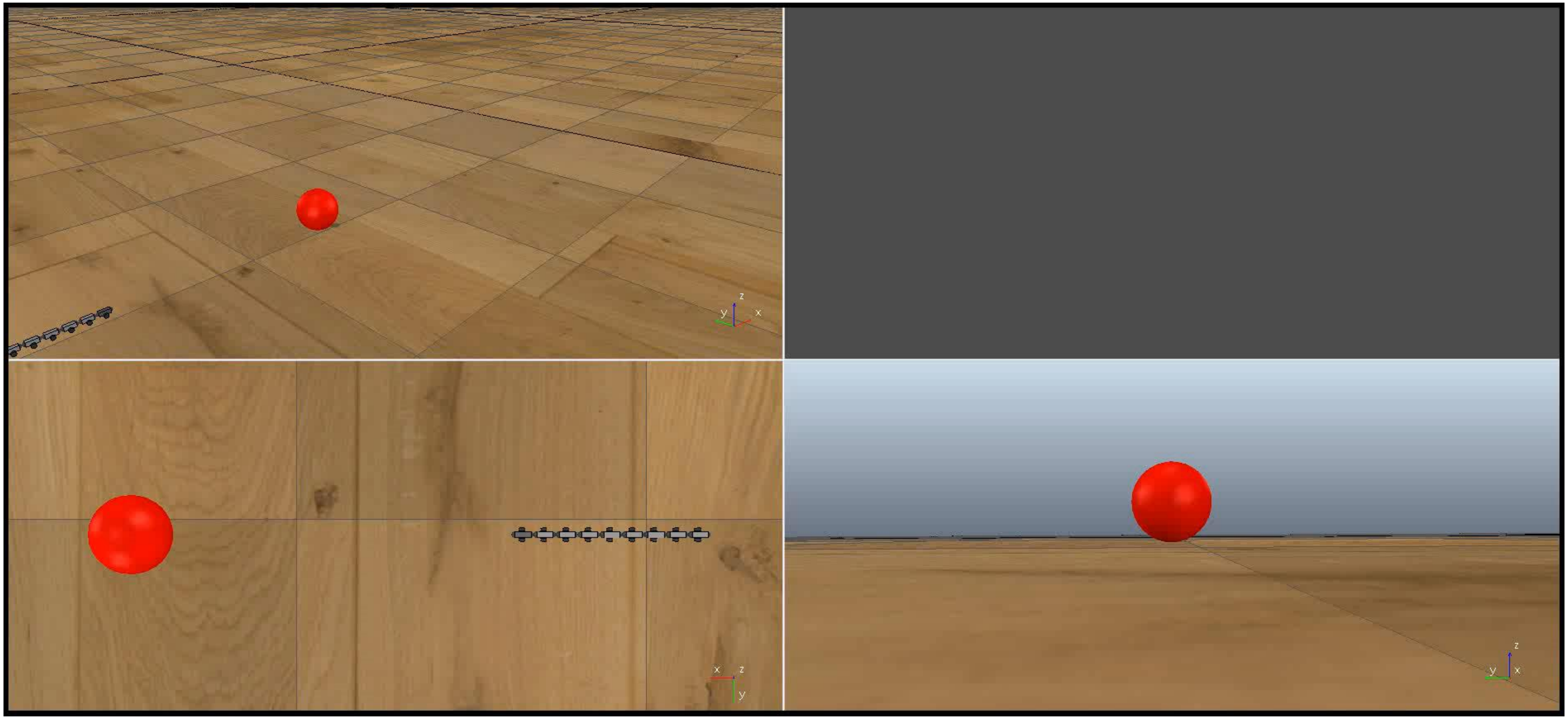




Results – Demonstration Indirect Agent



Results – Demonstration Direct Agent

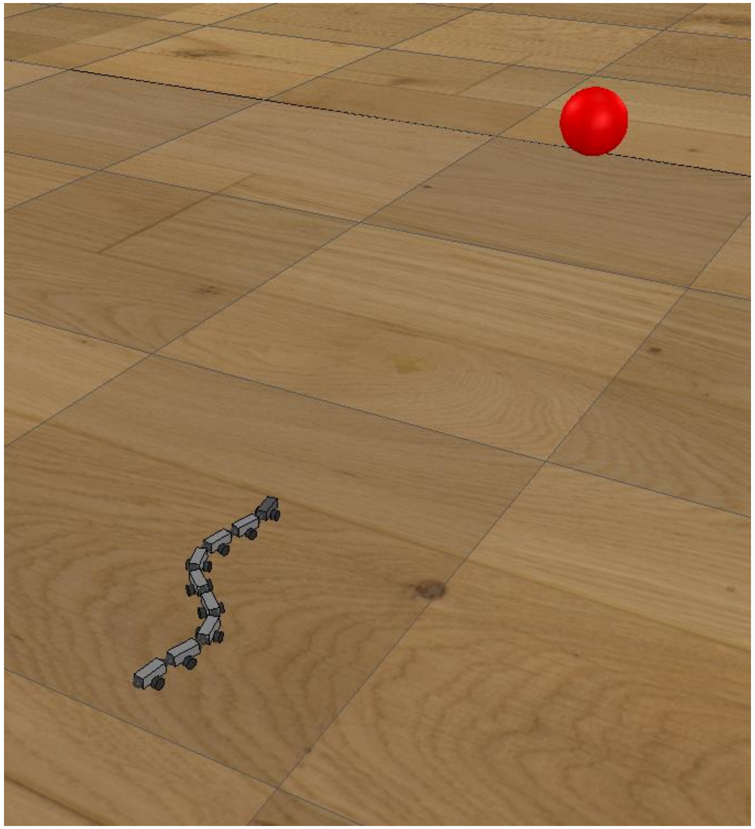


Agenda

- Motivation
- Methodology
- Approach
- Results
- Conclusion

Conclusion

- Both approaches successfully tracked the target in the training scenario
- Direct agent achieved higher target tracking accuracy and robustness
- Indirect agent performed unstable movement
- Indirect is more transparent and human operators can intervene



Future work:

- Proposal: RL locomotion, human steering
- Control all parameters of the slithering gait
- Rich environments with obstacles
- Applications for snake-like robots with 3D locomotion

Thank you for your attention

References

https://blog.openai.com/learning-dexterity/

https://wayve.ai/blog/learning-to-drive-in-a-day-with-reinforcement-learning

http://biorobotics.ri.cmu.edu/projects/modsnake/pictures.html

https://biorob.epfl.ch/salamandra

[1] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. 7 2017. URL http://arxiv.org/abs/1707.06347.

[2] Shigeo Hirose. Biologically inspired robots : snake-like locomotors and manipulators. Oxford University Press, 1993. ISBN 0198562616.

PPO Resources

PPO Paper: https://arxiv.org/abs/1707.06347 [1]

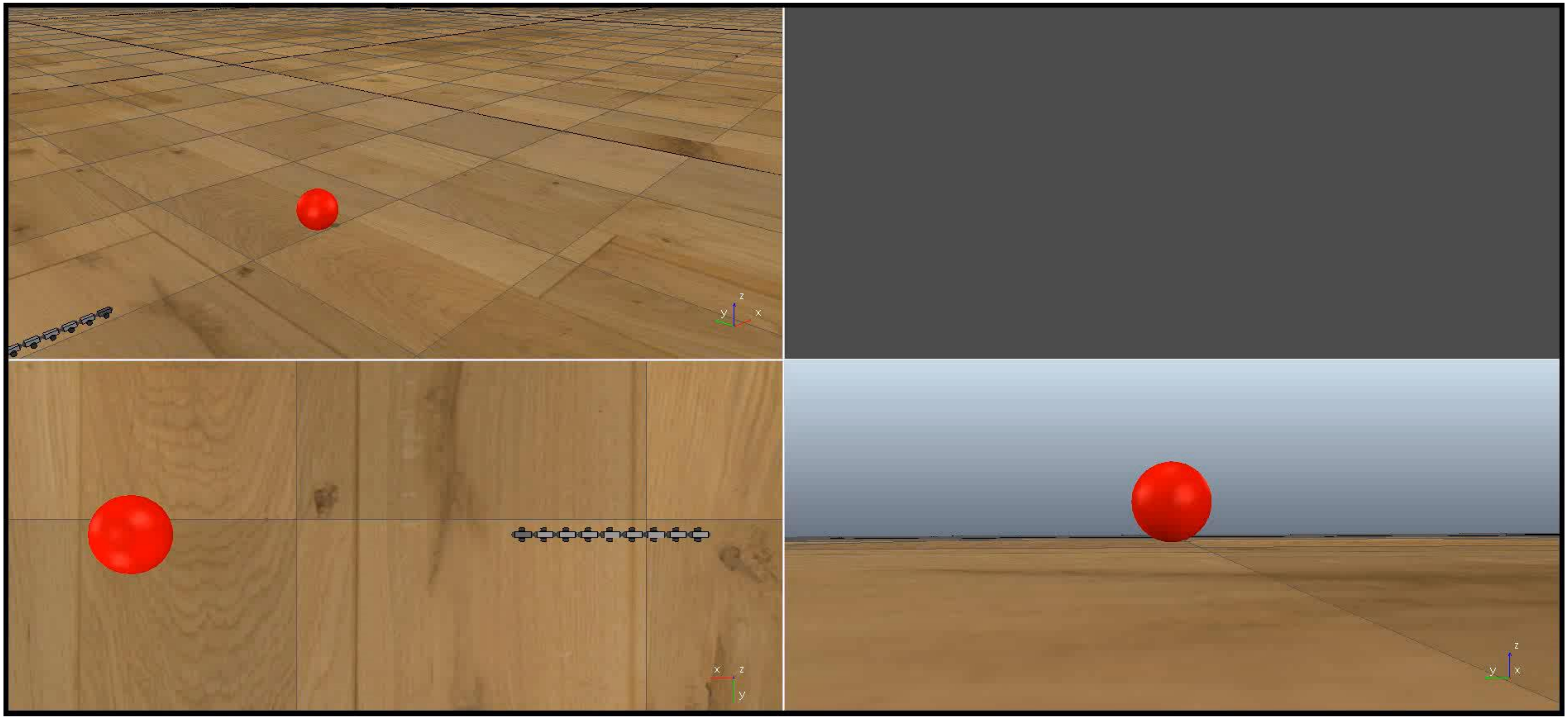
TRPO Paper: https://arxiv.org/abs/1502.05477

Arxiv Insights: https://www.youtube.com/watch?v=5P7I-xPq8u8

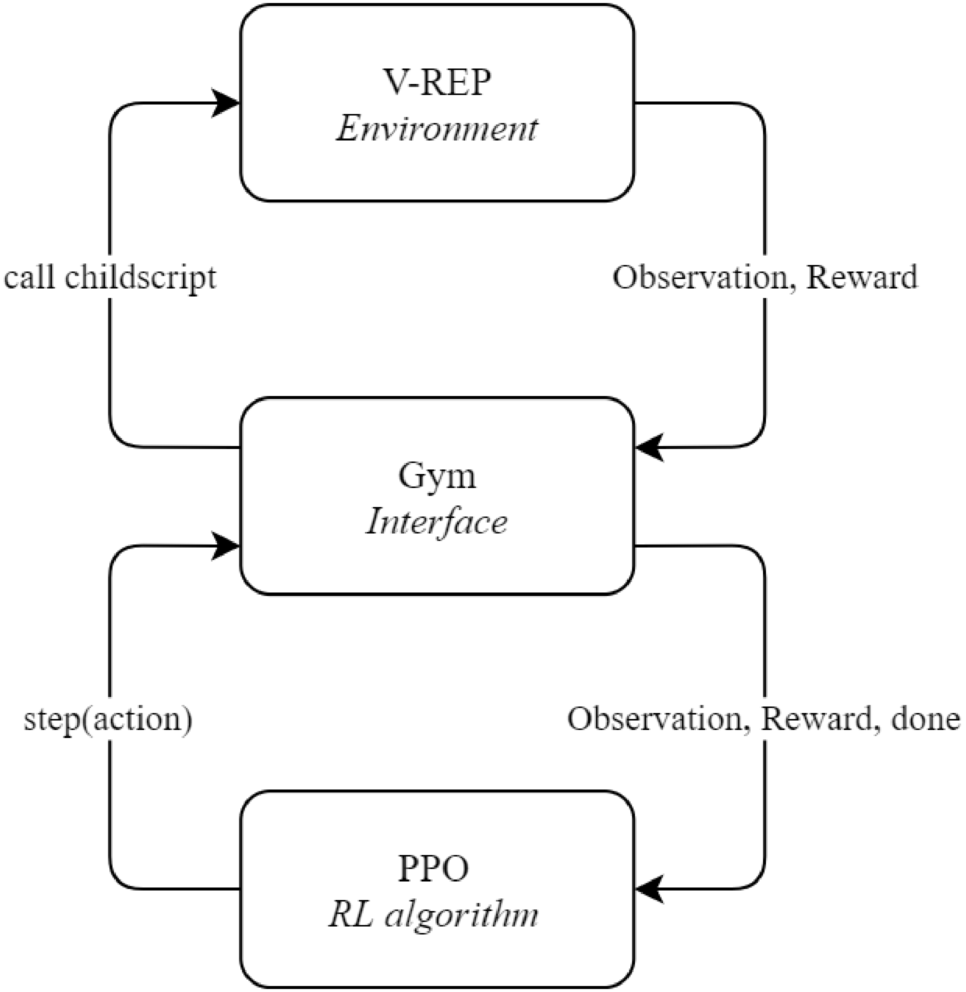
OpenAI blog: https://blog.openai.com/openai-baselines-ppo/

Deep RL Bootcamp - Lecture 5: https://youtu.be/xvRrgxcpaHY

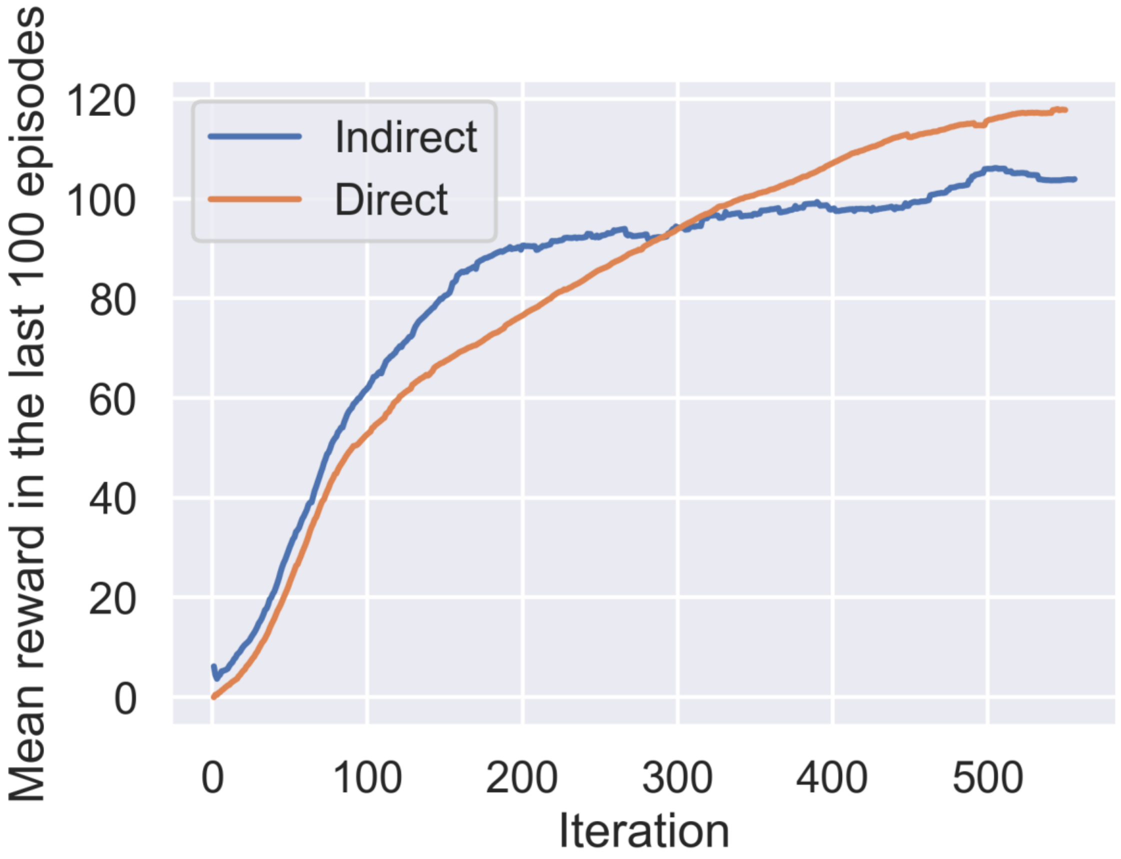
Direct Agent 1st Episode



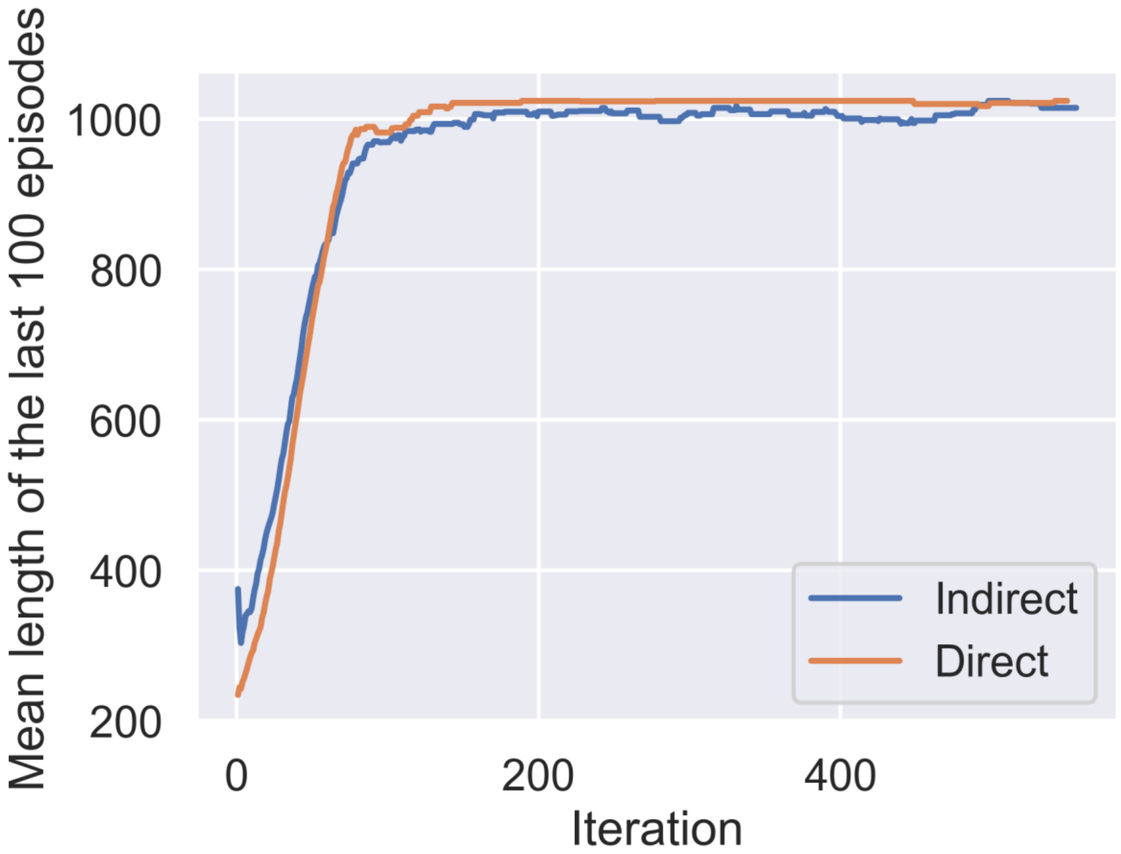
Methodology – Communication Overview



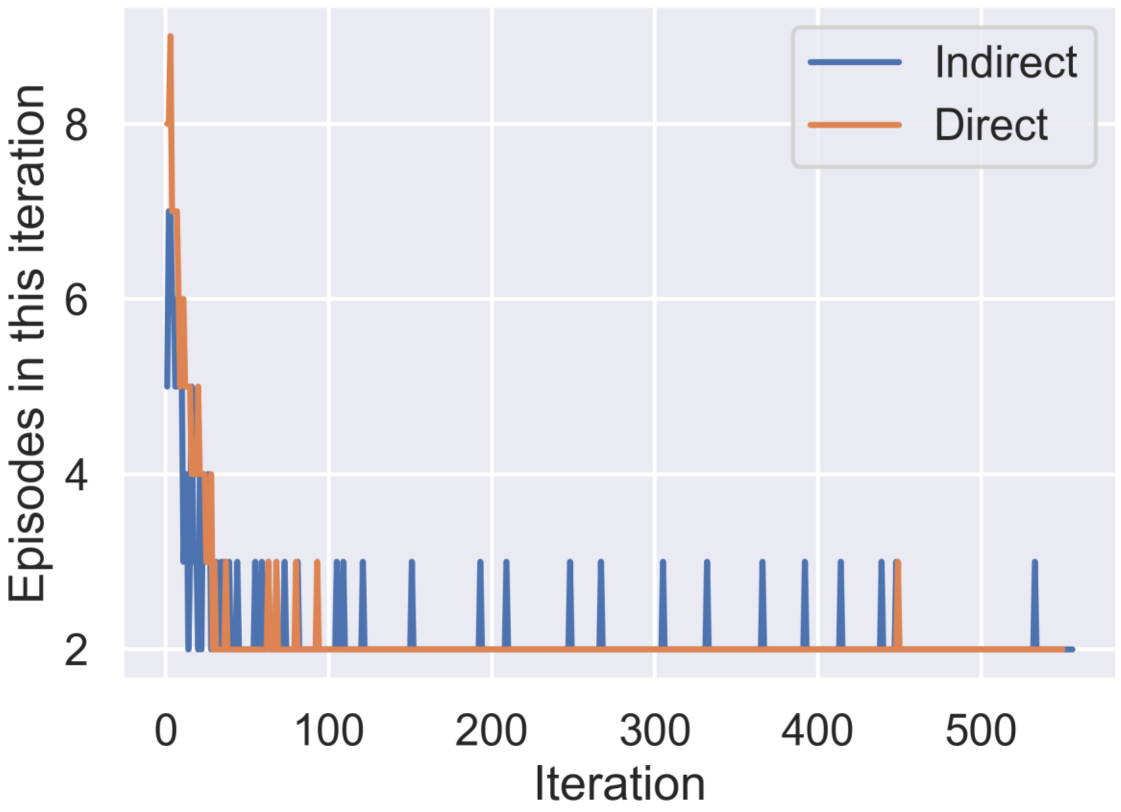
Training – Mean Reward



Training – Mean Episode Length



Training – Episodes per Iteration



Results - Comparison