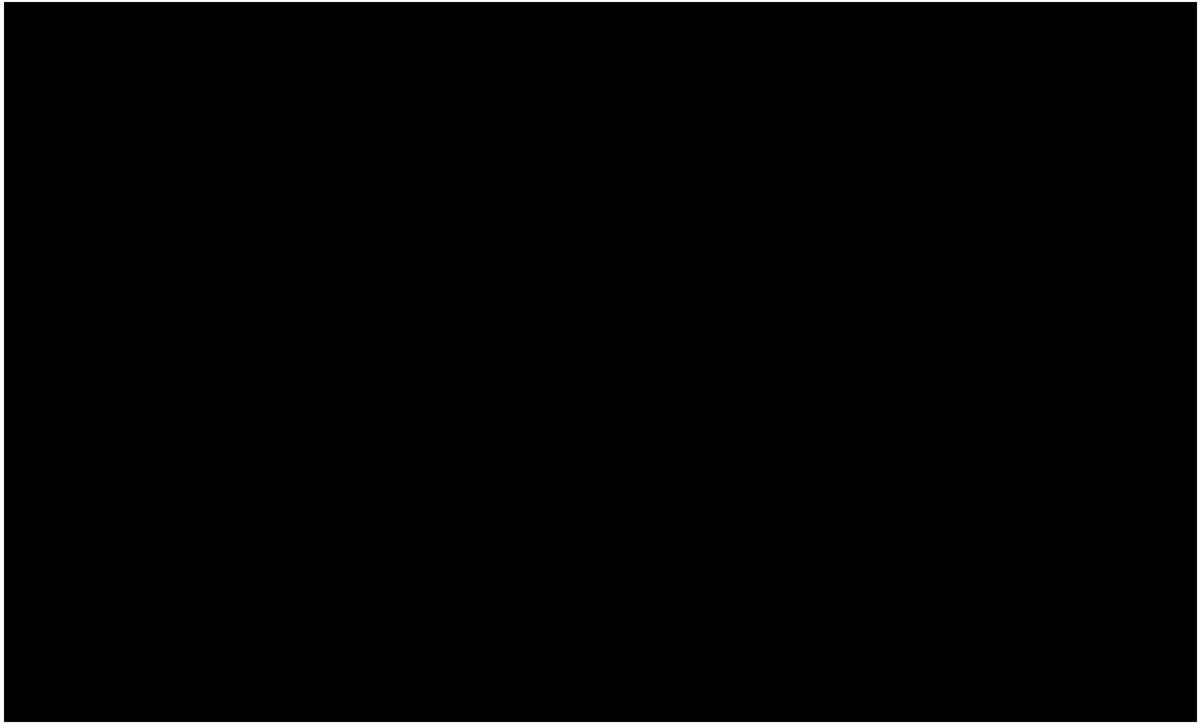
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Florian Vahl

Active Vision for Humanoid Soccer Robots using Reinforcement Learning



Motivation

- An accurate representation of the environment is required
- Only partial observations are possible

- Current approaches:
    - Keyframe animations
    - Object tracking

- New approach:
    - Learn a policy that controls the visual observations based on world model data

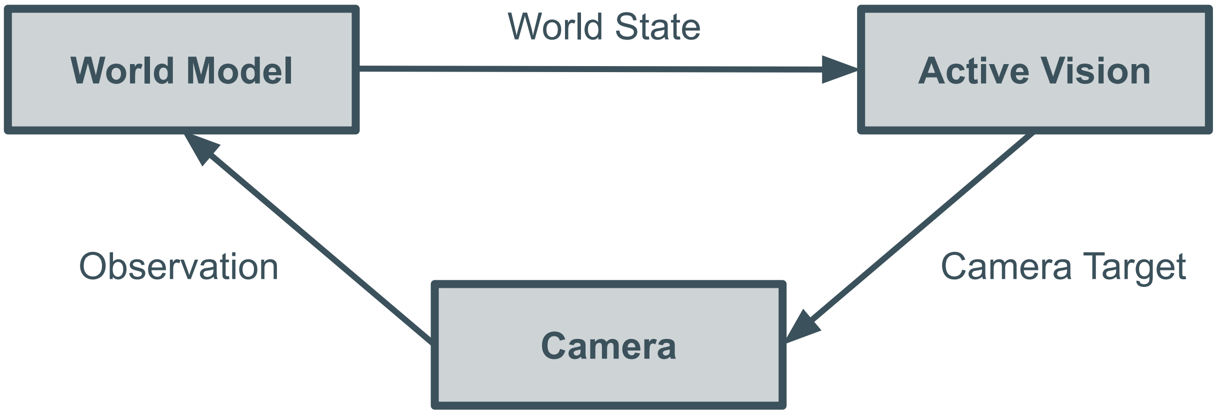
Active Vision

"We don't just see, we look"

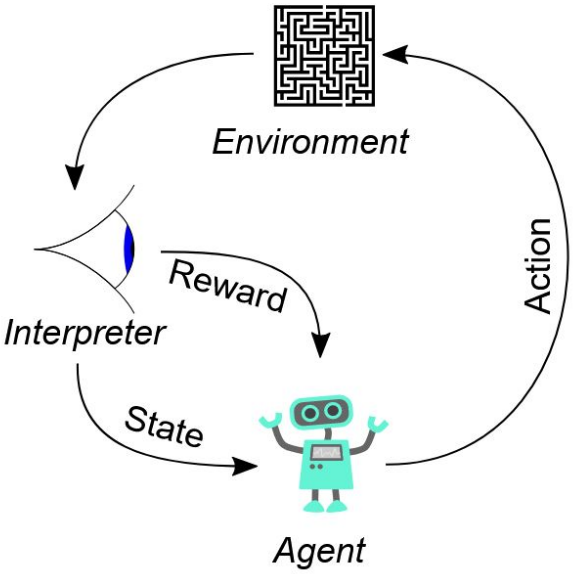
Ruzena Bajcsy in "Active perception"

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Active Vision

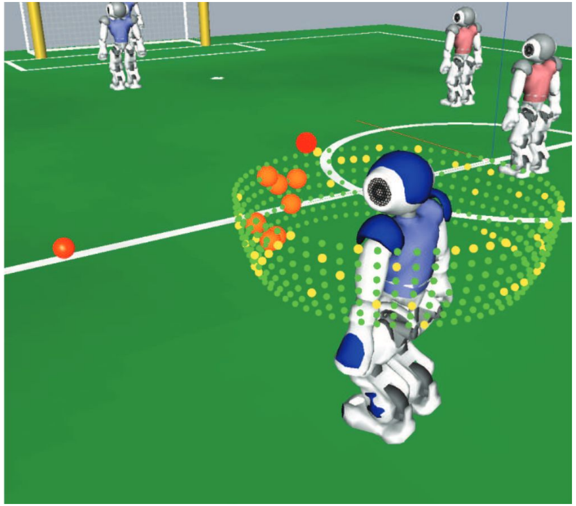


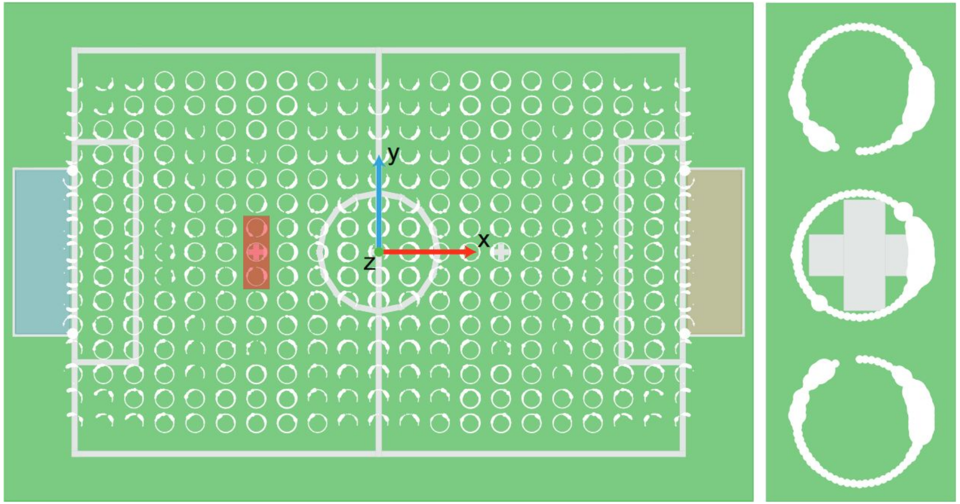
Reinforcement Learning



Source: Wikimedia Commons, CC-0

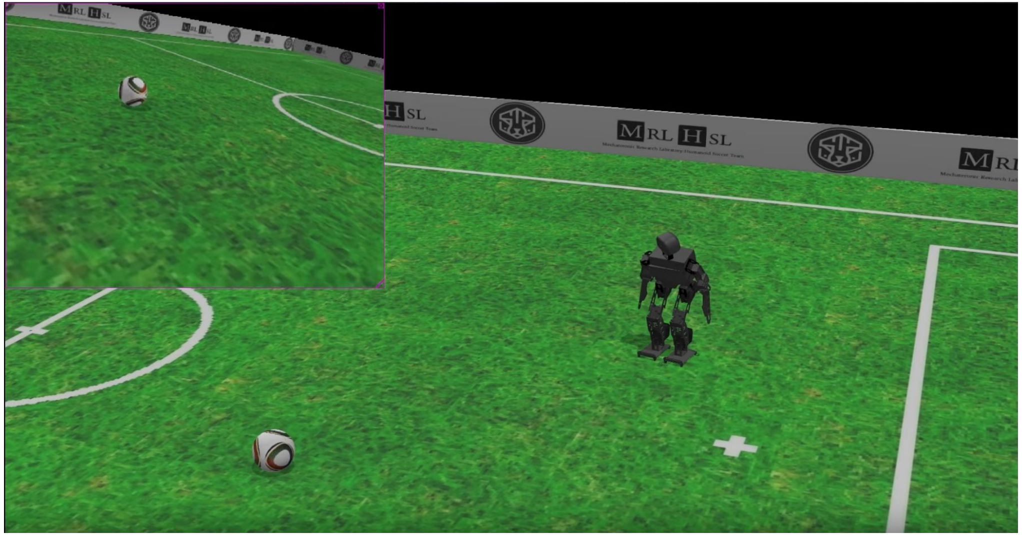
Entropy-Based Active Vision for a Humanoid Soccer Robot (Matías Mattamala et al., 2015)





Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

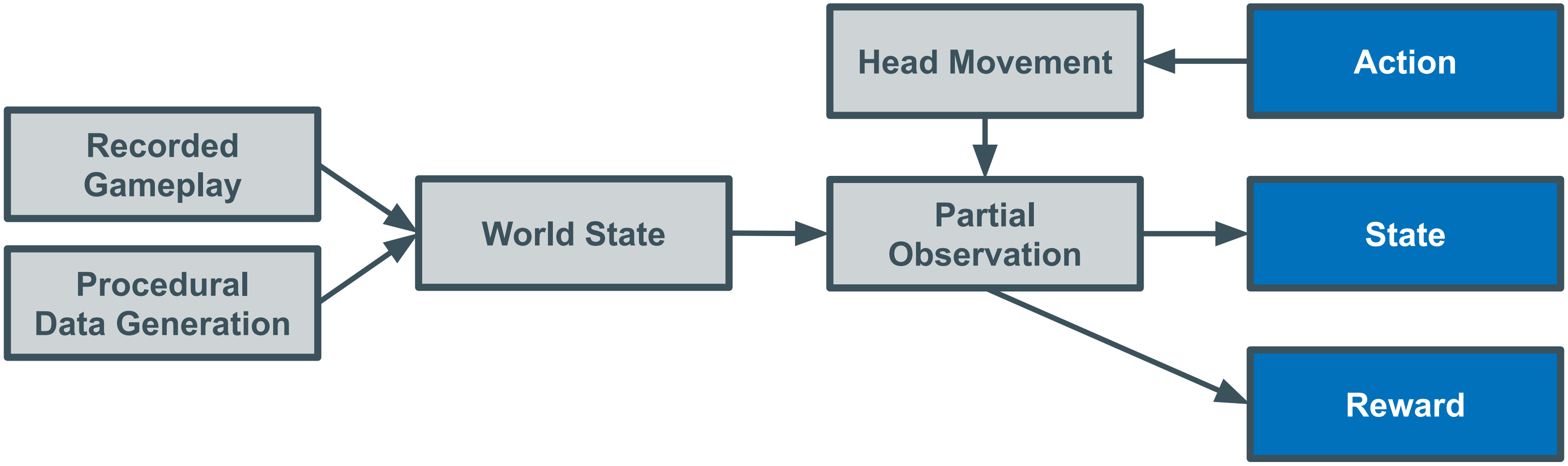
Real-time Active Vision for a Humanoid Soccer Robot Using Deep Reinforcement Learning (Soheil Khatibi et al., 2020)



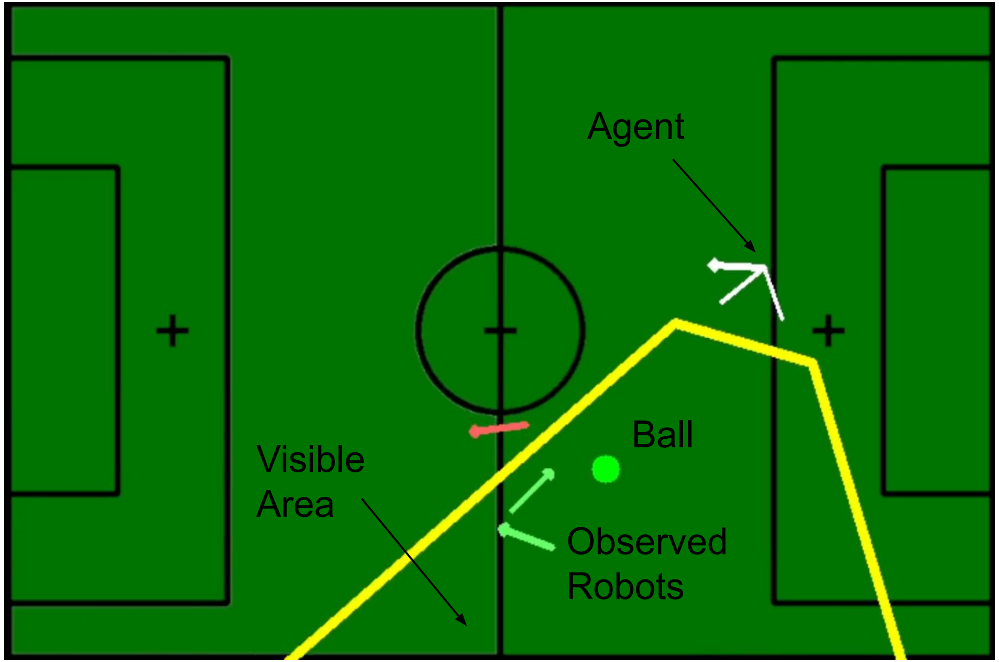
Training Environment

- Requirements:
  - Runtime efficient
  - Simulation of high level world model information
    - Robot / ball positions and visibility
    - Partial observation based on camera pose, projections and world state
  - Independent from specific team strategy / behavior
  - No detailed physics

Training Environment



Training Environment Visualization

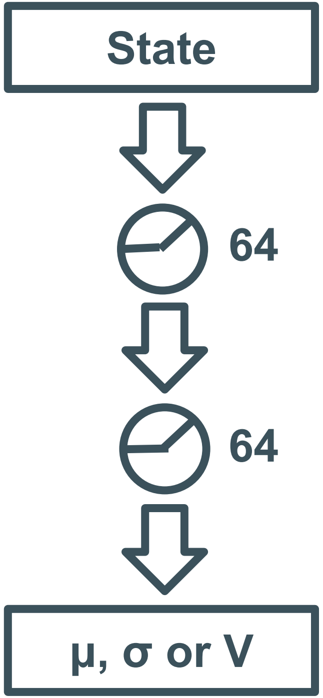


Training Setup

- Environment:
  - High-Level Simulator
  - Webots Simulator
  - Wolfgang Robot
  - Data from RoboCup 2021 and Brazil Open 2021
- Training:
  - Stable Baselines 3
  - Proximal Policy Optimization (or Deep Q-Learning)

Network Architecture

- Separate networks for policy and value function
- Two fully connected hidden layers with 64 neurons
- ReLU activation function
- Gaussian sampling with learned variance
- CNN encoder to process feature maps (optional)



Reward

- World Model Reward
  - Ball discrete visibility
  - Ball confidence
  - Robot discrete visibility
  - Robot confidence
  - Field coverage
- Imitation Reward
  - Sinusoidal demonstration MSE

$$r = \sum_{p \in R_{\mathrm{wm}}} w_p \cdot p + w_{\mathrm{demo}} \cdot R_{\mathrm{demo}} + R_{\mathrm{base}}$$

Actions

- Controls the neck joints in a continuous manner
- Different possibilities:
  - Cartesian actions
    - Image center projected onto the field (IK)
  - Joint actions
    - Joint velocities
    - Joint positions
    - Pattern (phase and amplitude of a sine function)

States

- Base footprint pose
- Camera pose
- Neck joint positions
- Neck joint position history
- Phase
- Action history
- Ball position and confidence
- Robot positions and confidences
- Feature maps
  - Robot position map
  - Viewed field regions as Motion History Image

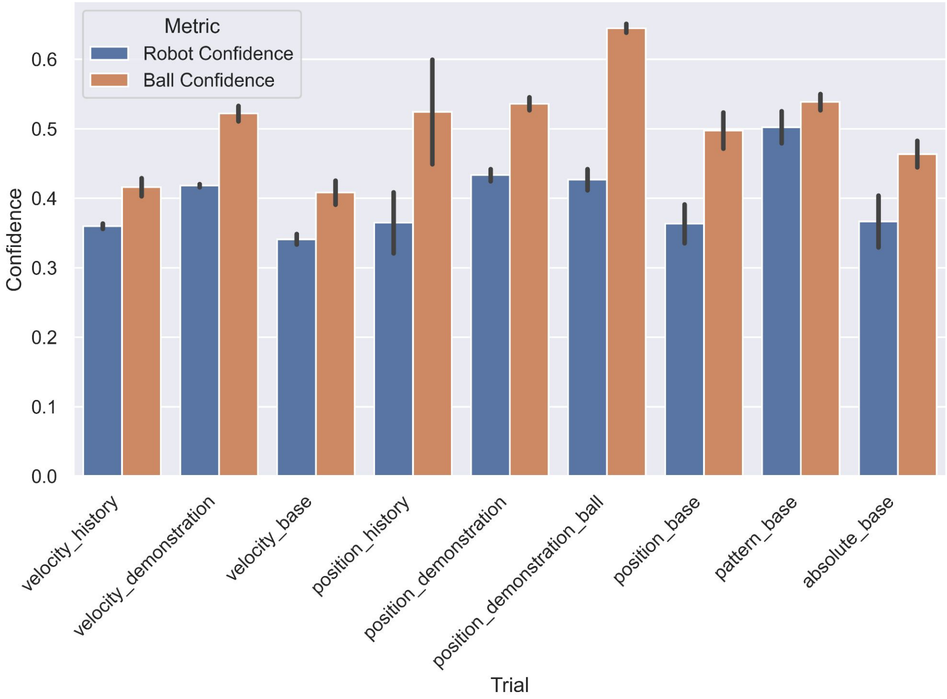
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Combinations

Name	Neck joints	Base footprint	Camera pose	Ball state	Robot states	Phase	Neck joints history	Action history
velocity_history	x	x	x	x	x		x	x
velocity_demo	x	x	x	x	x	x		
velocity_base	x	x	x	x	x			
position_history	x	x	x	x	x		x	x
position_demo	x	x	x	x	x	x		
position_demo_ball	x	x	x	x		x		
position_base	x	x	x	x	x			
pattern_base	x	x	x	x	x			
abs_base	x	x	x	x	x			

Experiments

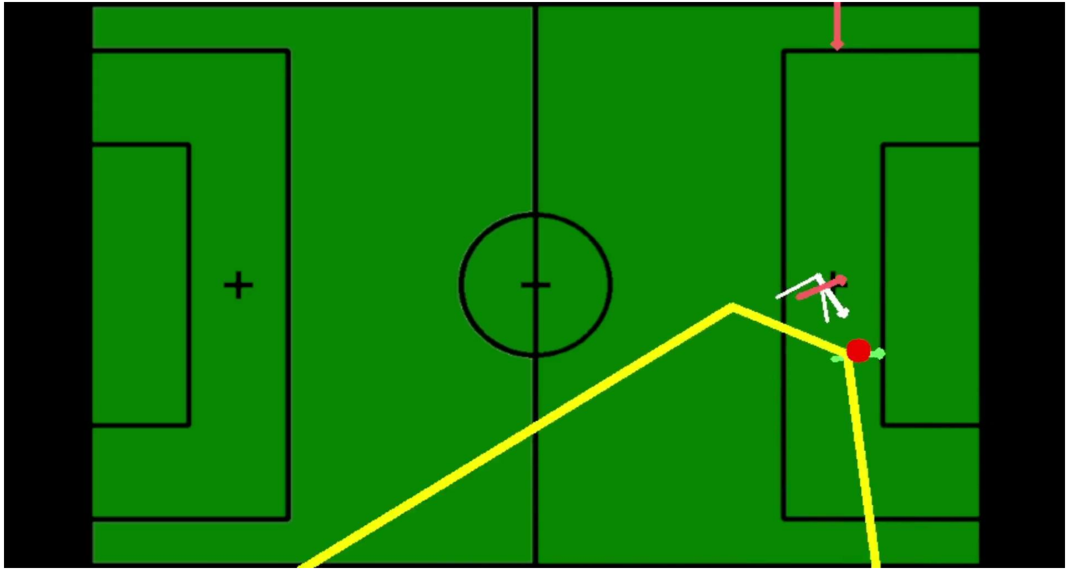
Experiments



Qualitative Examples

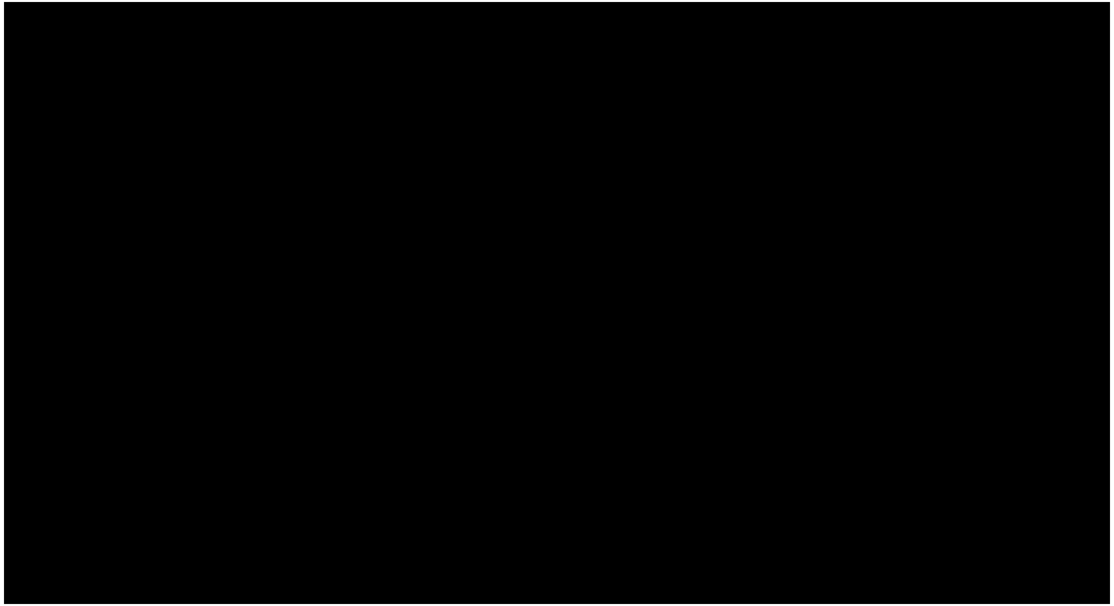
Policy trained with demonstration

Qualitative Examples



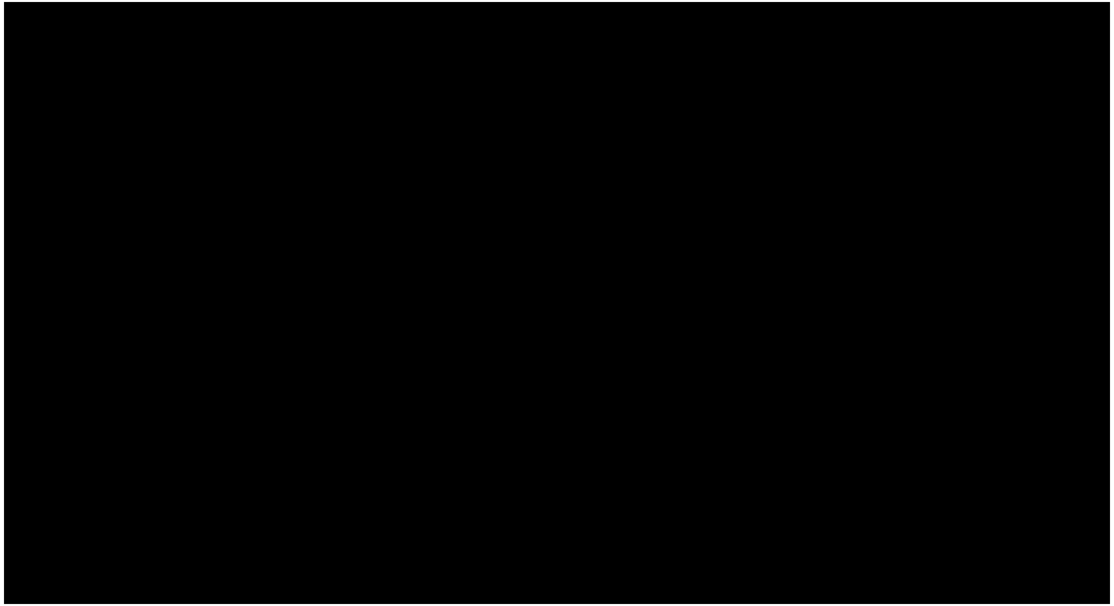
Policy trained with demonstration

Qualitative Examples



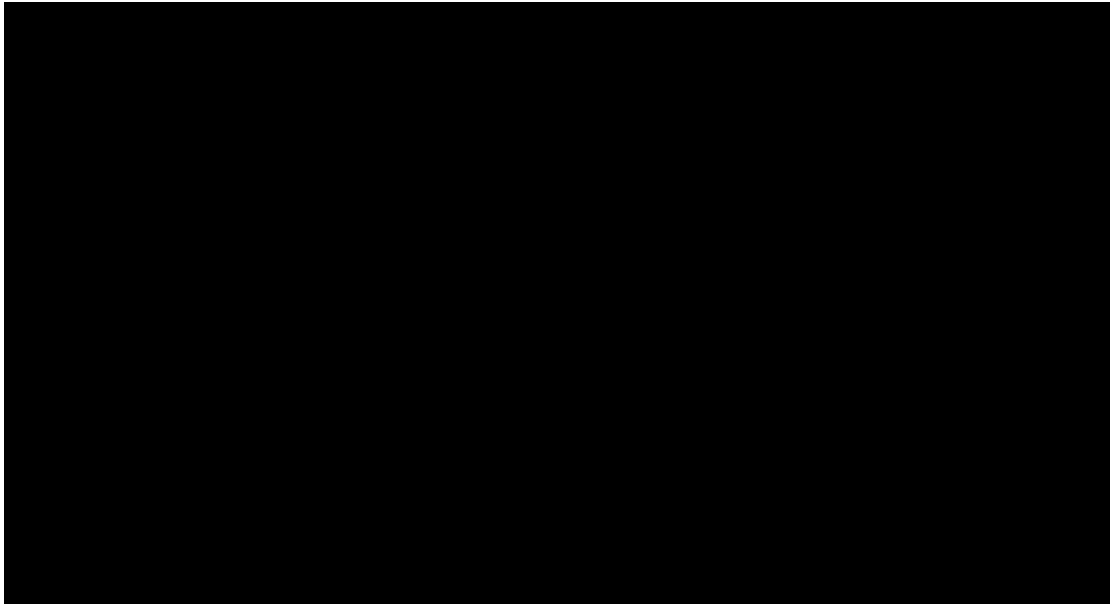
Policy trained without history or demonstration

Qualitative Examples



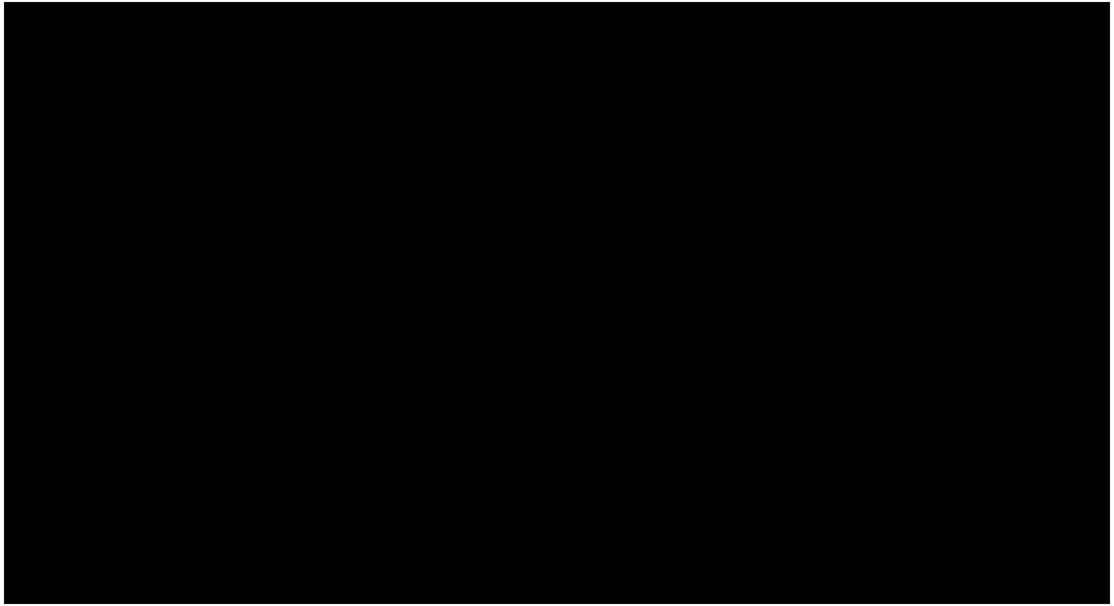
Policy trained with action and state history

Qualitative Examples



Policy with pattern actions

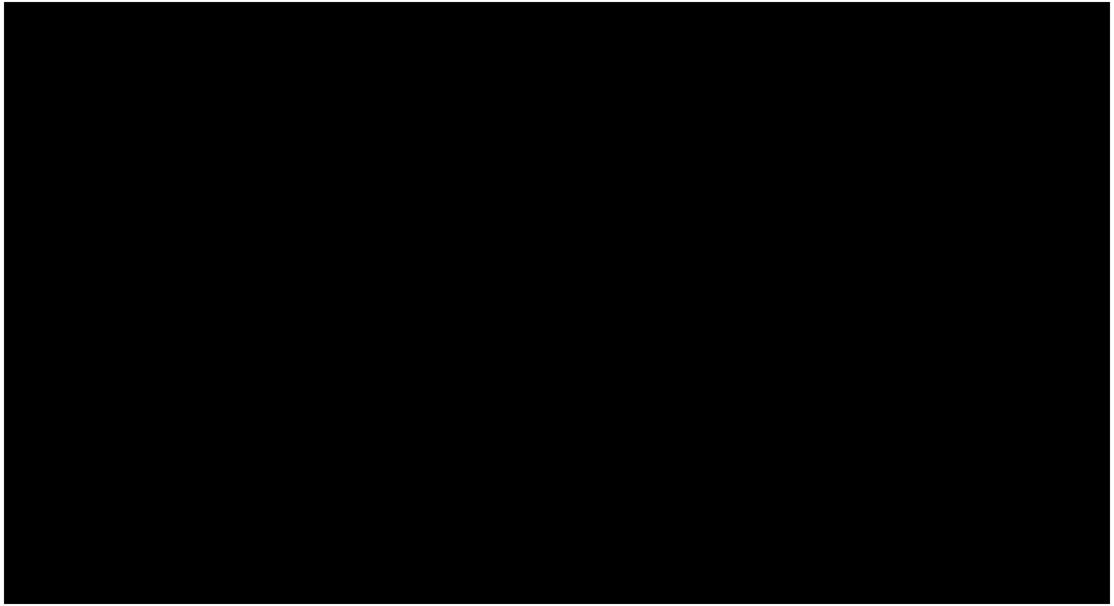
Qualitative Examples



Policy with absolute position action (IK)

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Webots



Trained policy running alongside the Bit-Bots software stack

Results and further Work

- Training pipeline works as intended
- Demonstration as well as history increased performance
- Joint positions and pattern seem to be the best action spaces

- Further work
  - Evaluate feature map observations
  - Evaluate other hyperparameters (MLP size, History length, …)

Questions?

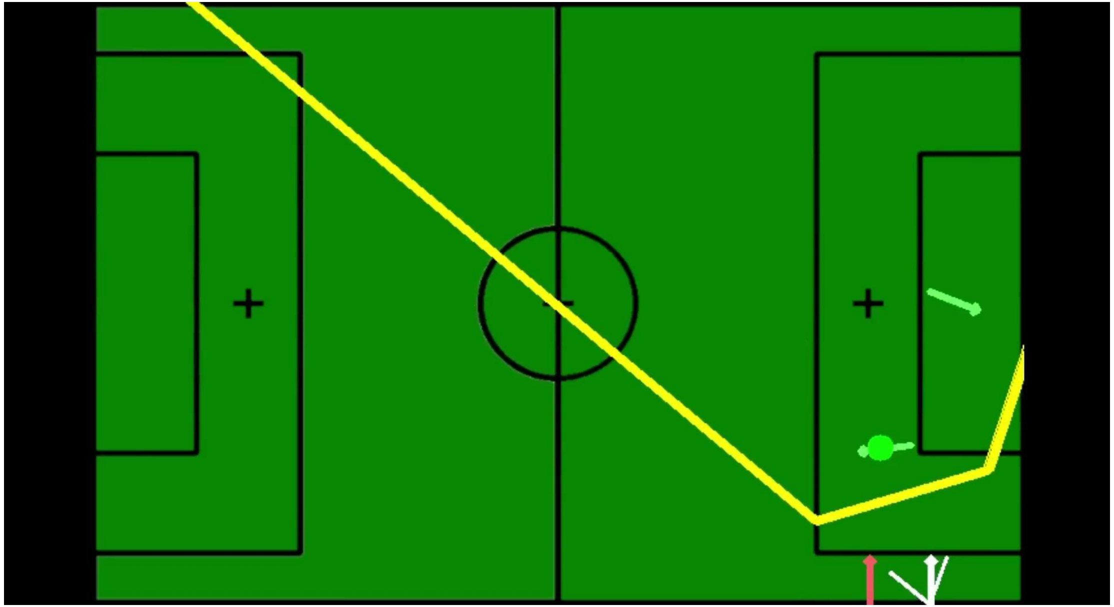
Appendix

Motion History Image

Robot Platform

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Qualitative Examples



Policy trained with demonstration