Directed Curriculum Generation for Reinforcement Learning

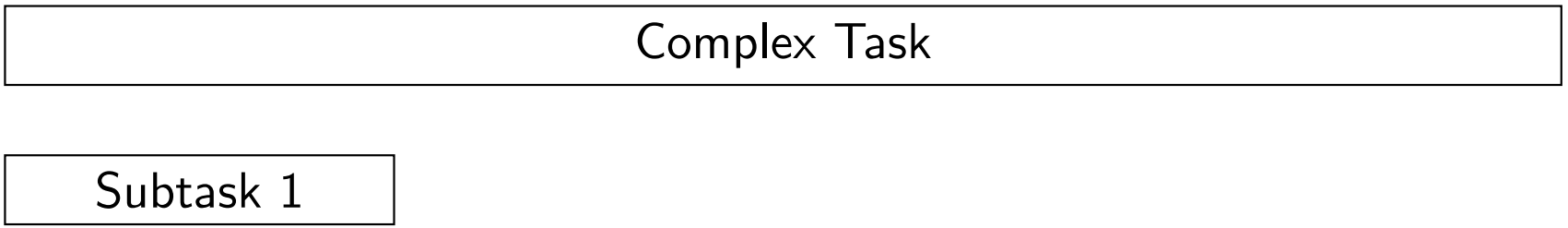
Thomas Edlich

September 19, 2018

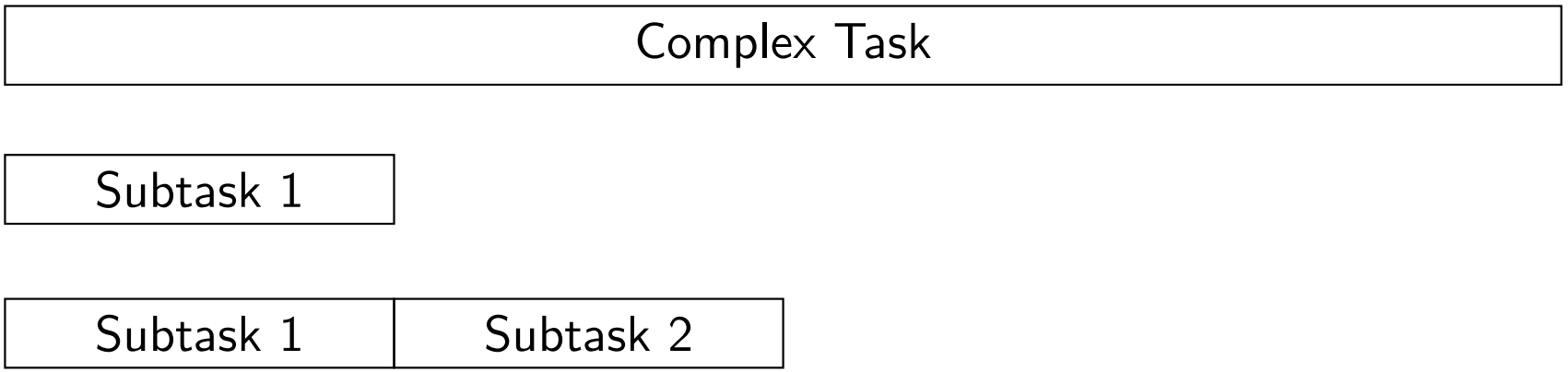
Curriculum Learning

Complex Task

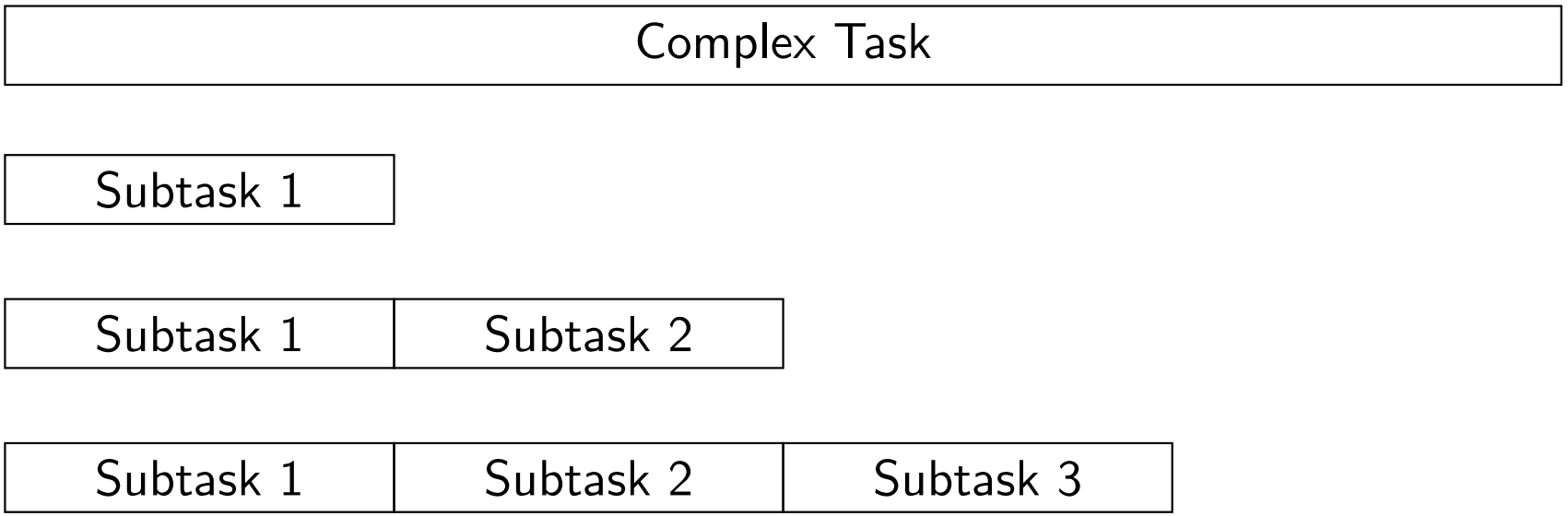
Curriculum Learning



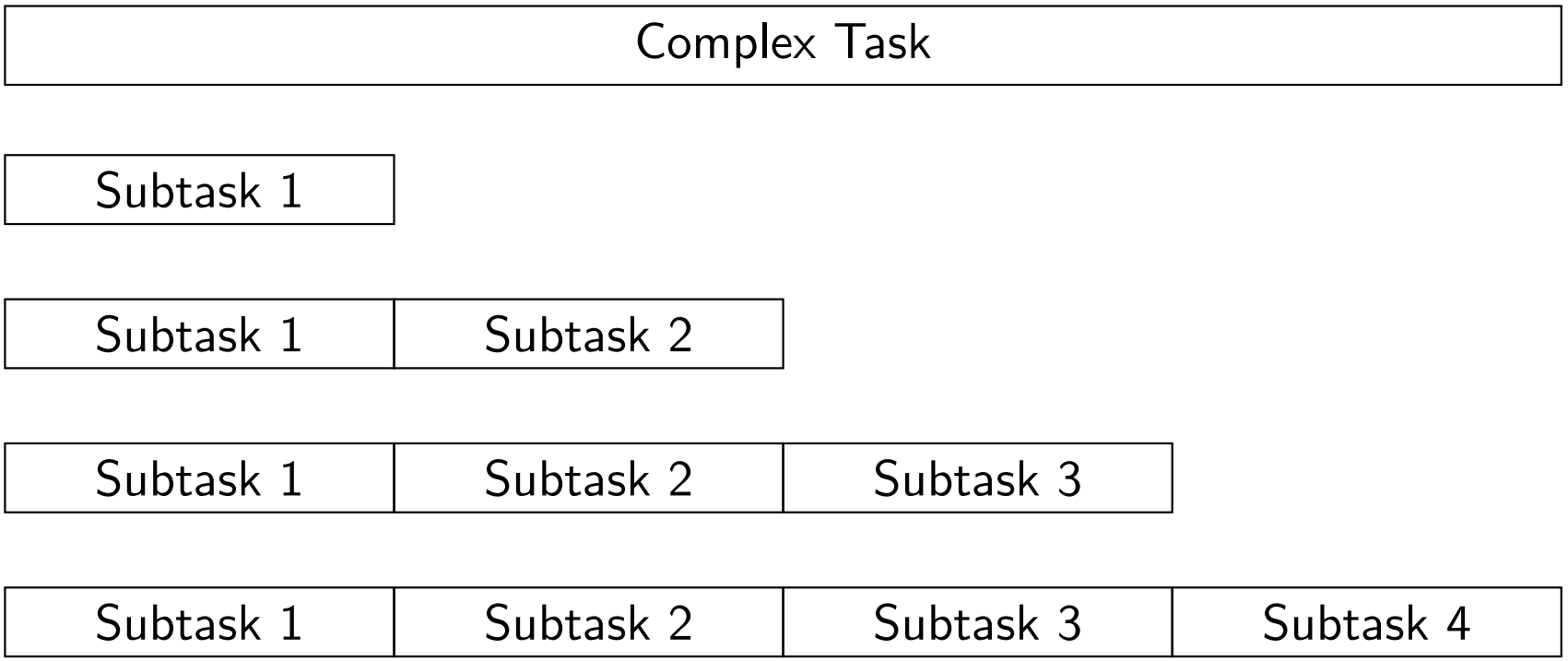
Curriculum Learning



Curriculum Learning



Curriculum Learning



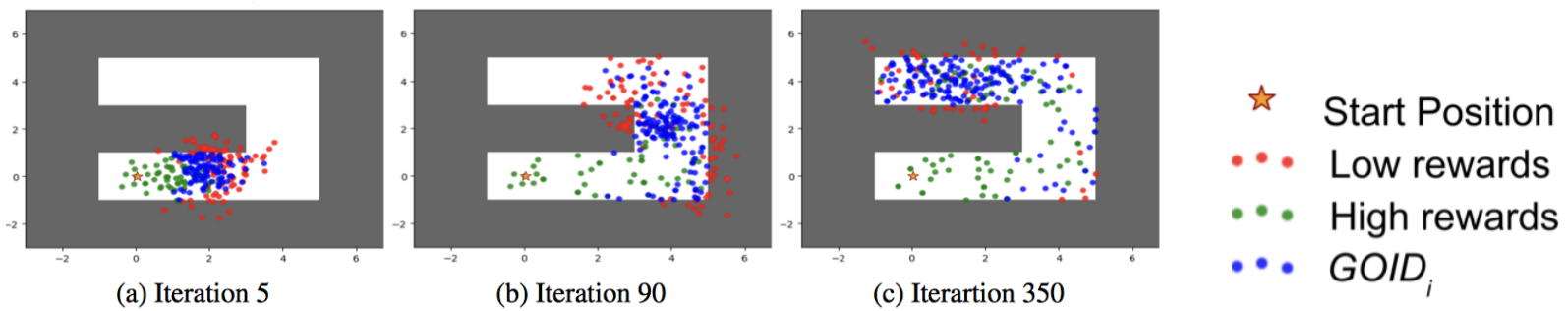
Curriculum Learning in Reinforcement Learning

- Curricula often relying on manually defined stages, e.g. Karpathy and Van De Panne 2012
- Recent approaches to automatic curriculum generation:
  - Reverse Curriculum Generation Florensa et al. 2017
  - Region-Growing Curriculum Generation: Molchanov et al. 2018
  - Intrinsic Motivation and Automatic Curricula via Asymmetric Self-Play
  - Automatic Goal Generation for RL Agents (GoalGAN) Held et al. 2018
  - Hindsight Experience Replay Andrychowicz et al. 2017

Automatic Curriculum Generation Approaches I

Automatic Goal Generation for RL Agents (GoalGAN)[1]

- Task: from one starting point learn to a policy to reach any feasible state $s \in \mathcal{S}$ by creating a curriculum of increasingly difficult goals
- Use a GAN to generate *Goals of Intermediate Difficulty* (GOID), i.e. goals which can be reached by the current policy ($\mathbb{E}[R] > R_{min}$) but are not too easy ($\mathbb{E}[R] < R_{max}$)
- Goal Replay Buffer to prevent forgetting



(a) Iteration 5 (b) Iteration 90 (c) Iterartion 350

Figure 1: Examples of generated Goals by GoalGAN.

[1] Held et al. 2018

Automatic Goal Generation for RL Agents (GoalGAN)[2]

Algorithm 1 Automatic Goal Generation For RL

$G, D \leftarrow pretrain\_GAN()$
$goals_buffer \leftarrow \emptyset$
for $i = 1, \ldots$ **do**
    $goals \leftarrow G(z) \cup sample(goal_{replaybuffer})$
    $\pi_i \leftarrow update\_policy(goals, \pi_{i-1})$
    $labels \leftarrow label\_goals(goals, \pi_i)$
    $G, D \leftarrow train\_GAN(goals, labels)$
    $goal_{replaybuffer} \leftarrow store\_goals(goals, goals_{replaybuffer})$
end for

[2] Held et al. 2018

Motivation

Directed Curriculum Learning

- Curriculum generation algorithms expand based on difficulty of task
- None of these algorithms consider the usefulness of a task
- We aim to modify existing curriculum generation algorithms in order to expand mainly towards regions of high similarity with a pre-defined goal region

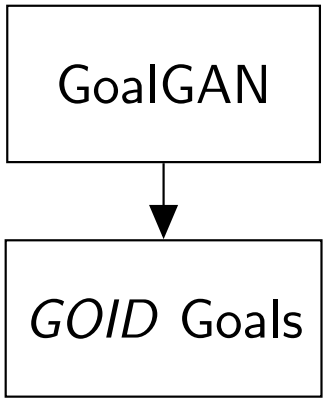
General Approach

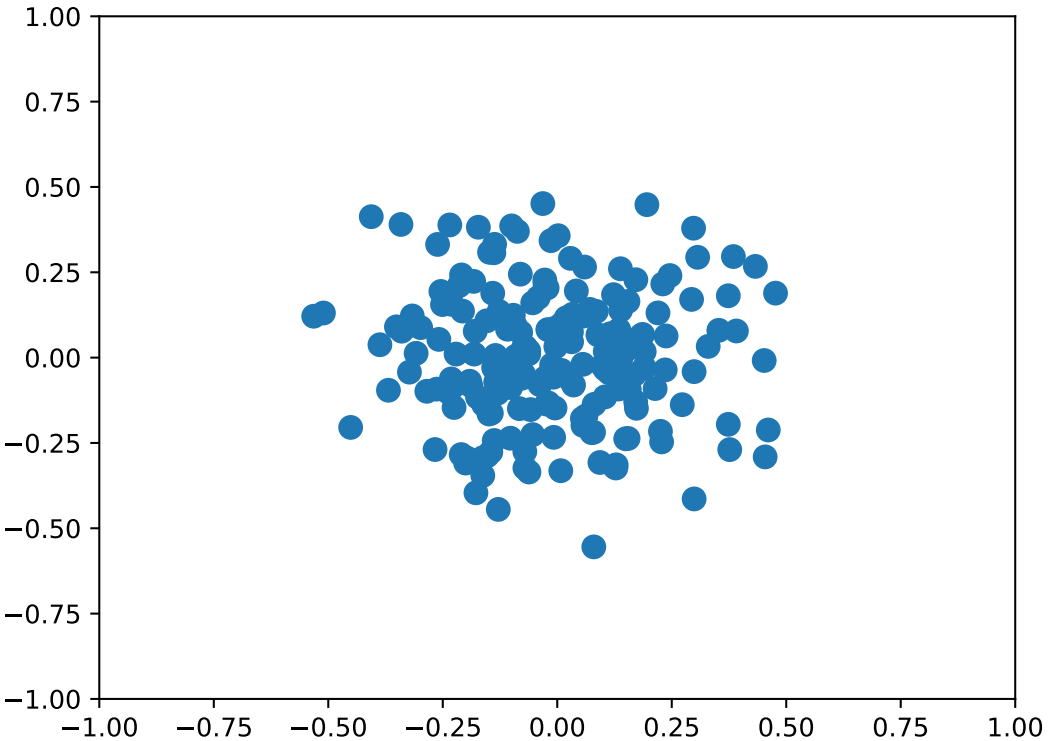
Directed Curriculum Generation

- Train a discriminator $\mathcal{D}$ to distinguish between target goals $s_g \in \mathcal{S}_g$ and so far achieved goals
- Discriminator can be used to rate the quality of a state, i.e. how useful/similar it is to the target goal states
- Idea: Train primarily on goals with higher discriminator ratings, then update discriminator
- Discriminator guides selected goals towards the provided target states
- Using the discriminator's ratings we can sample the best goals from a given set

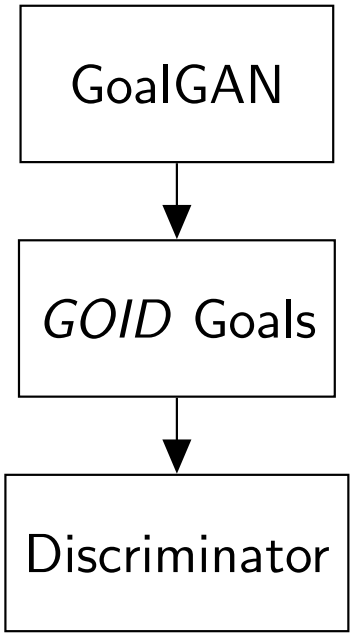
Directed GoalGAN

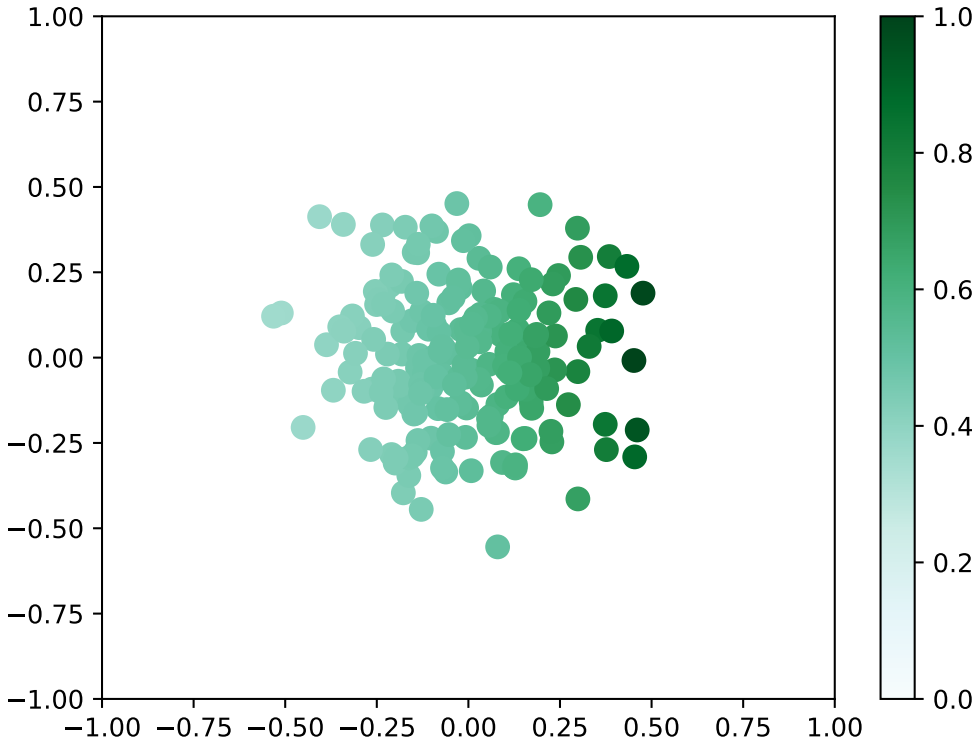
Directed GoalGAN



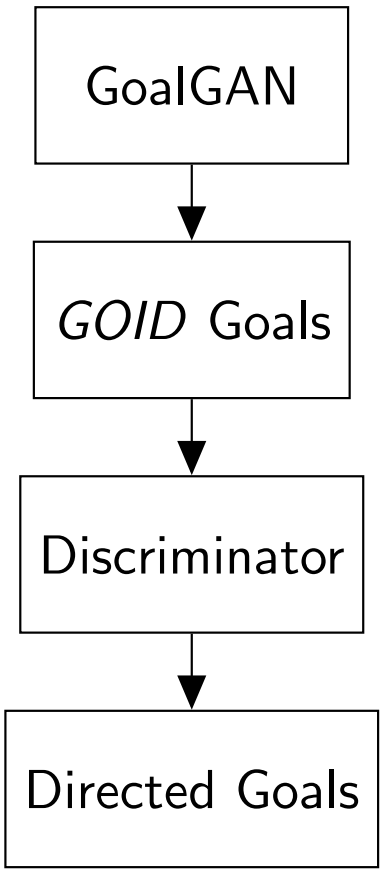


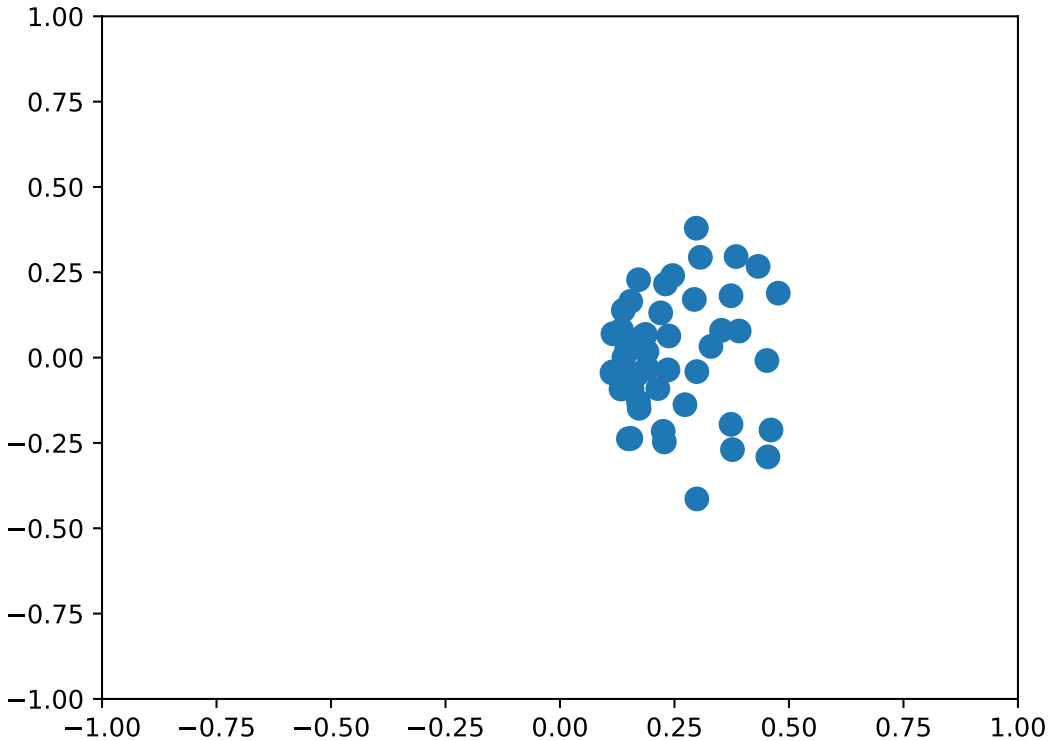
Directed GoalGAN





Directed GoalGAN





Experiment: GoalGAN vs Directed GoalGAN

- State Space: $(x, y, vel_x, vel_y, dist_x^+, dist_x^-, dist_y^+, dist_y^-)$
- Goal Space: $[-1, 1]^2$ (position)
- Action Space: $[-1, 1]^2$ (velocity adjustment)

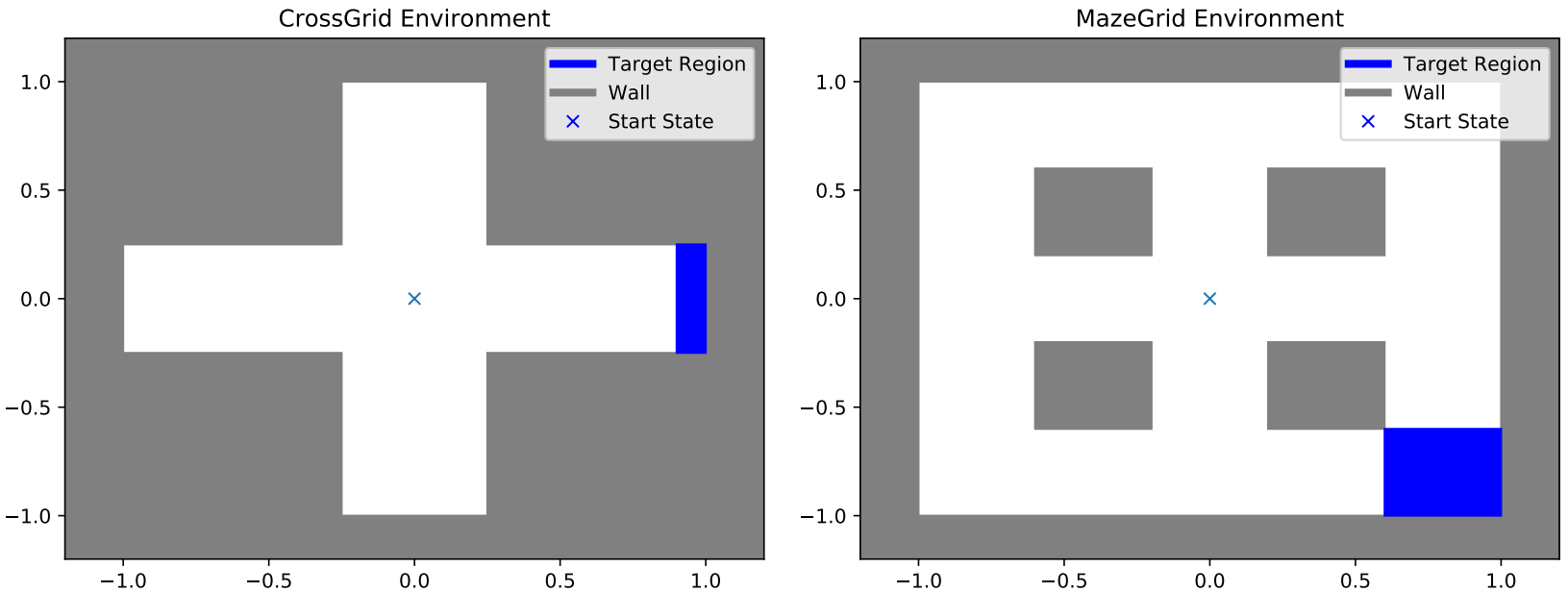


Figure 2: Layout two different grid world environments.

Experiments: GoalGAN vs Directed GoalGAN

Coverage Development of GoalGAN

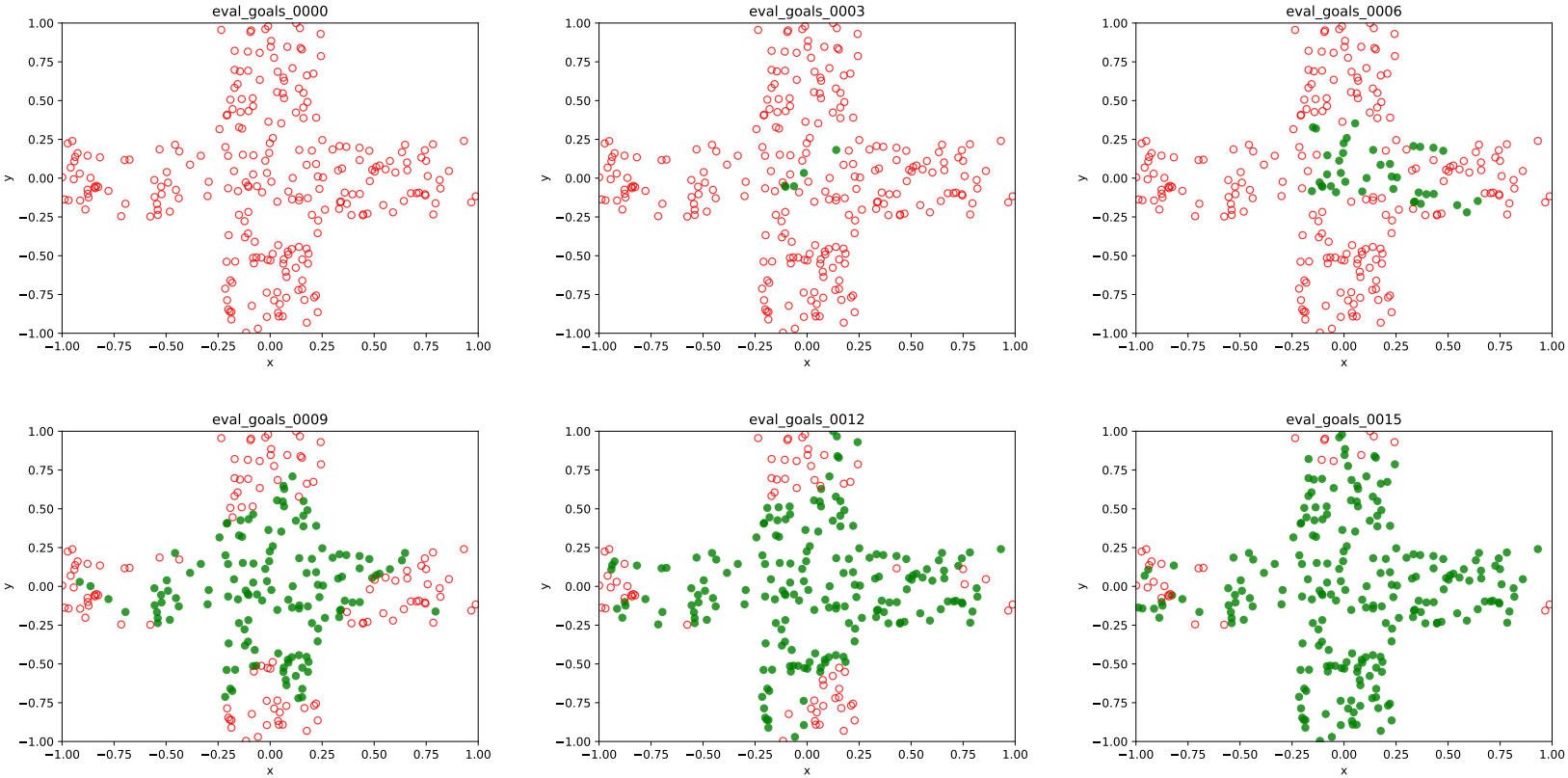


Figure 3: Coverage of the whole state space on the CrossGrid environment using GoalGAN. (Red=not reached, Green=reached)

Experiments: GoalGAN vs Directed GoalGAN

Coverage Development of Directed GoalGAN

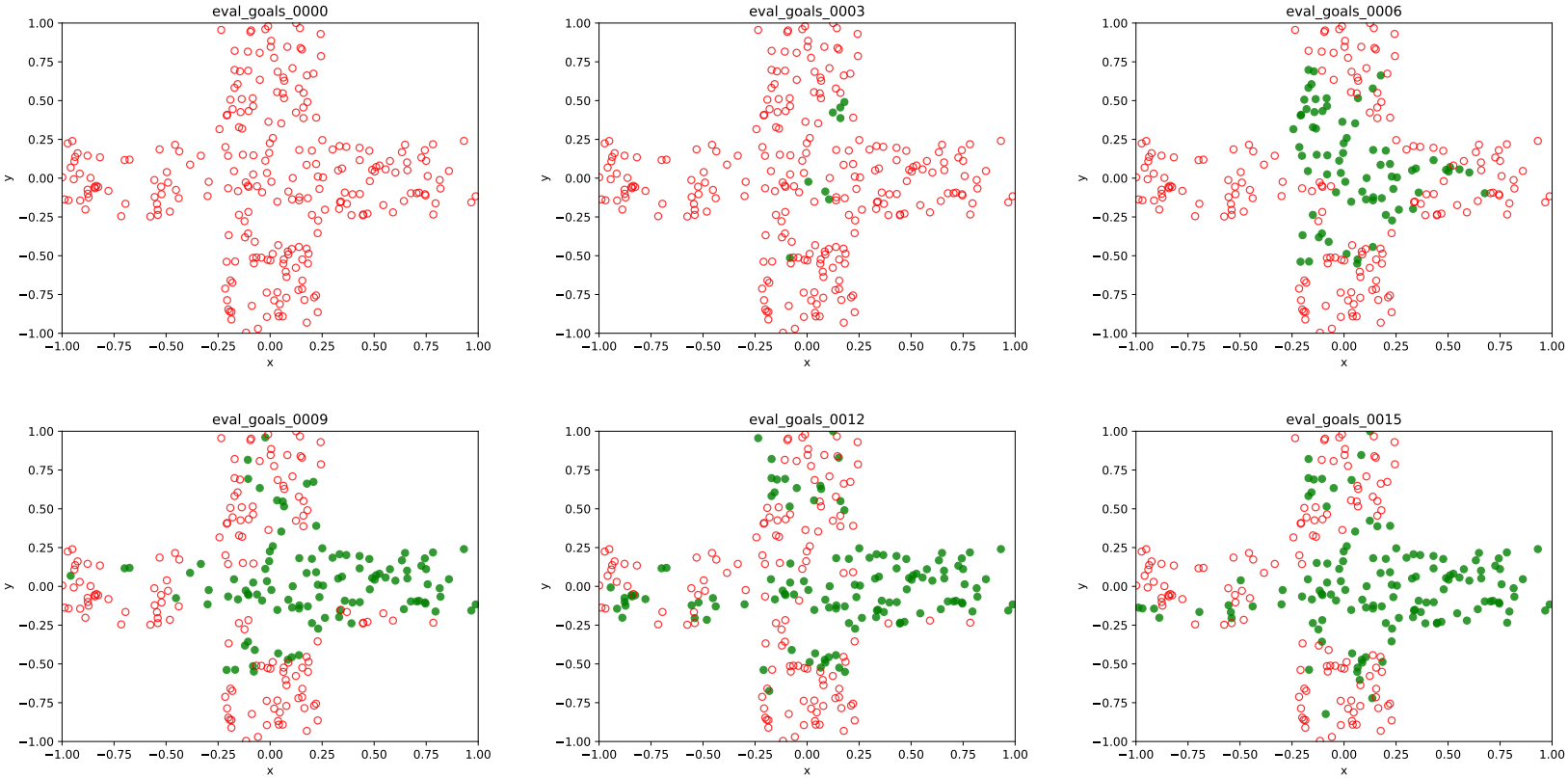


Figure 4: Coverage of the whole state space on the CrossGrid environment using GoalGAN. (Red=not reached, Green=reached)

Experiments: GoalGAN vs Directed GoalGAN

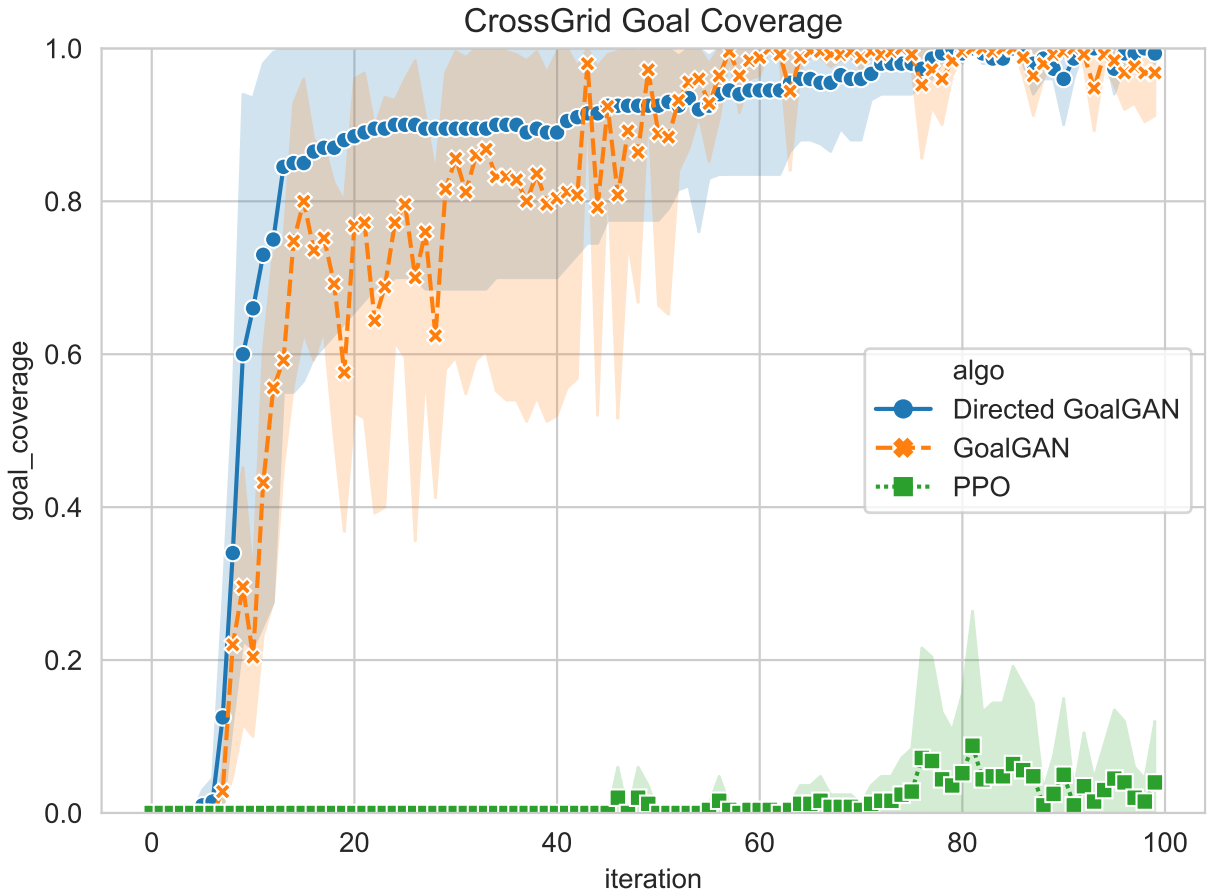


Figure 5: Coverage of the target region of the CrossGrid environment of Directed GoalGAN, GoalGAN, PPO (with uniform goal selection)

Experiments: GoalGAN vs Directed GoalGAN

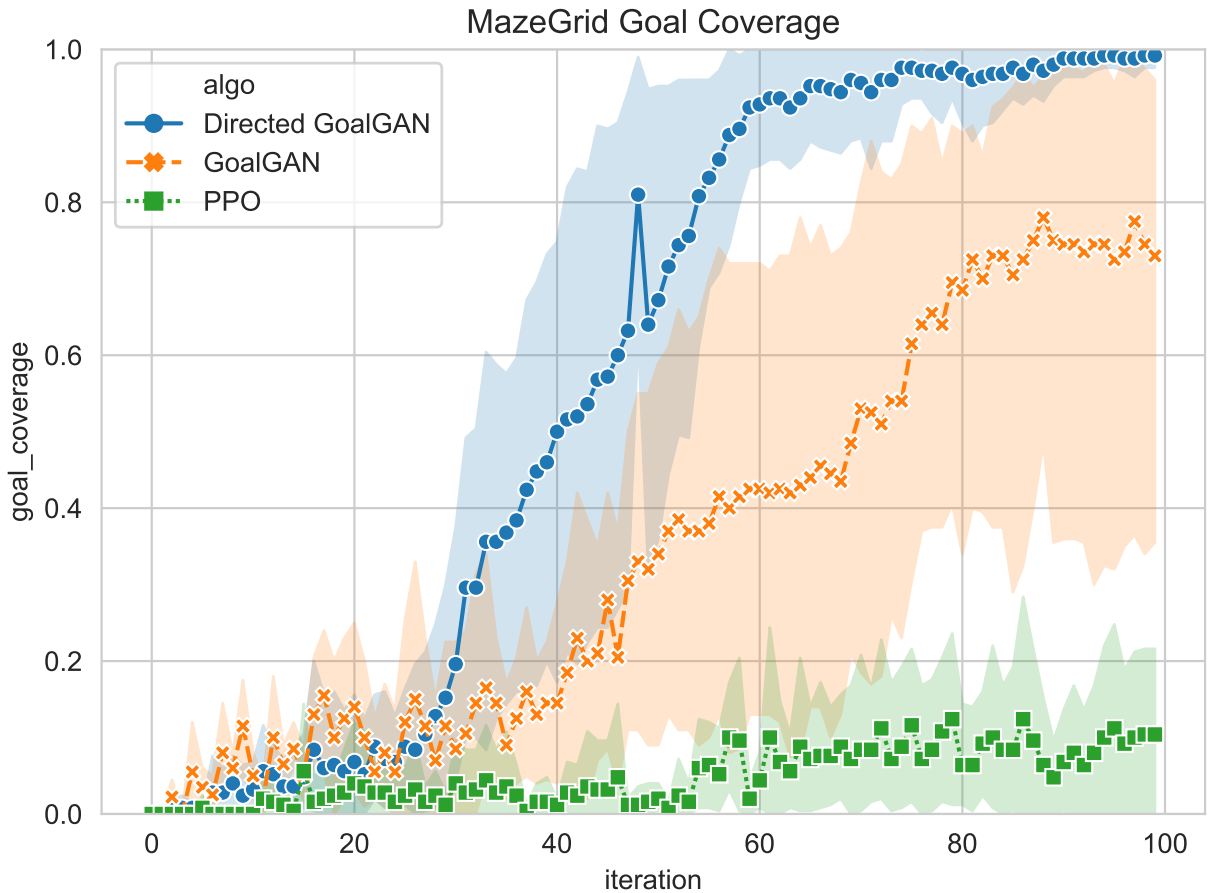


Figure 6: Coverage of the target region of the MazeGrid environment of Directed GoalGAN , GoalGAN, PPO (with uniform goal selection)

Experiments: Continuous GridWorlds Numerical Results Summary

- Directed GoalGAN reaches the target region faster than GoalGAN
- Directed GoalGAN covers less of the whole state space which is normal since we want to explore towards the goal region
- Problem I: GoalGAN approach is not sample-efficient
- Problem II: GoalGAN has problems in high dimensional goal spaces

Hindsight Experience Replay[3]

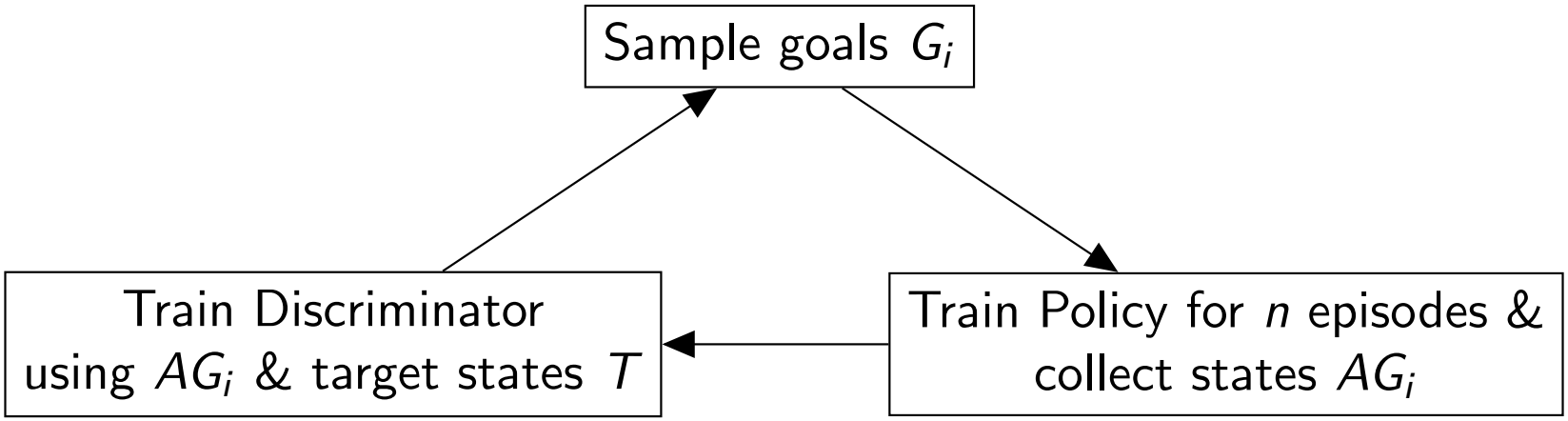
- Learning in hindsight by replacing the goal in collected experiences:

$$(s_t, a_t, r_t, s_{t+1}, g) \rightarrow (s_t, a_t, r_t, s_{t+1}, g')$$

- HER creates an implicit curriculum by learning even from failed attempts
- HER first learns to reach goals which the agent can achieve even by random behaviour (most likely states near the start state) and slowly learns to reach goals further and further away

[3] Andrychowicz et al. 2017

Directed HER



Intuition

Agent discovered something interesting (i.e. a state with a high discriminator rating), therefore set this state as a goal in order to explore this region more.

Directed HER

Algorithm 2 Directed HER

Require: Target States T, n_{goals}, $n_{targetgoals}$

$D \leftarrow init\_discriminator$
$goals = random\_sample(\mathcal{S}, n_{goals})$
for $i = 1, \ldots$ **do**
    $goals = goals \cup random\_sample(T, n_{targetgoals})$
    for $j = 1, \ldots n_{goals} + n_{targetgoals}$ **do**
        $initial\_goal = sample\_initial\_goal(goals, targets, \epsilon)$
        $achieved\_goals_j = HER(goals_j)$
        $scores_j = D.rate(achieved\_goals_j)$
        $new\_goal_j \leftarrow sample(achieved\_goals_j, scores_j, 1)$
    end for
    $goals = \{new\_goal_j\}$
    $D = train\_discriminator(achieved\_goals_i, T)$
end for

Experiment: Directed HER

Does Directed HER generate a directed curriculum?

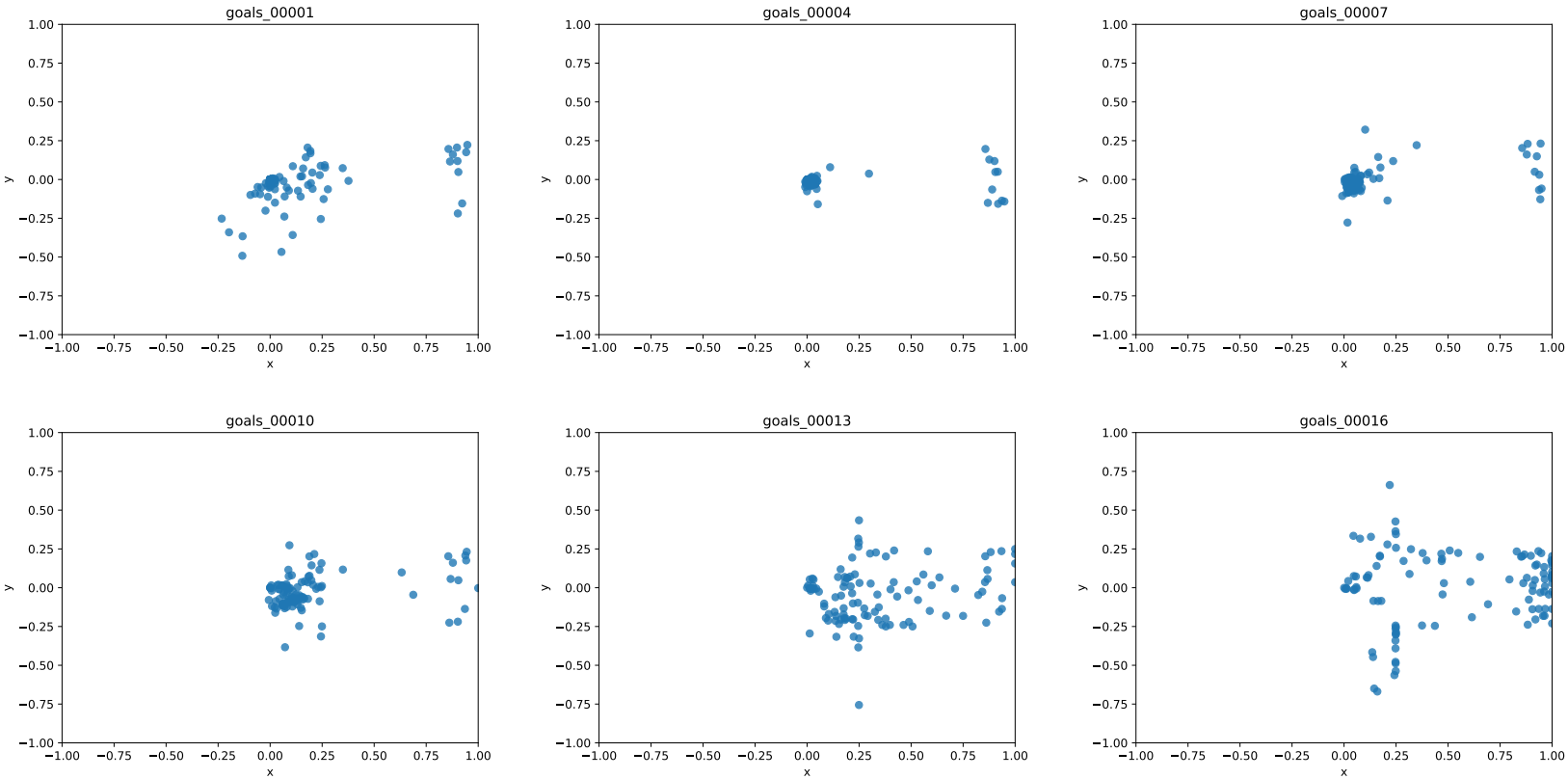


Figure 7: Initials goals selected by Directed HER over time.

Experiments: Directed HER

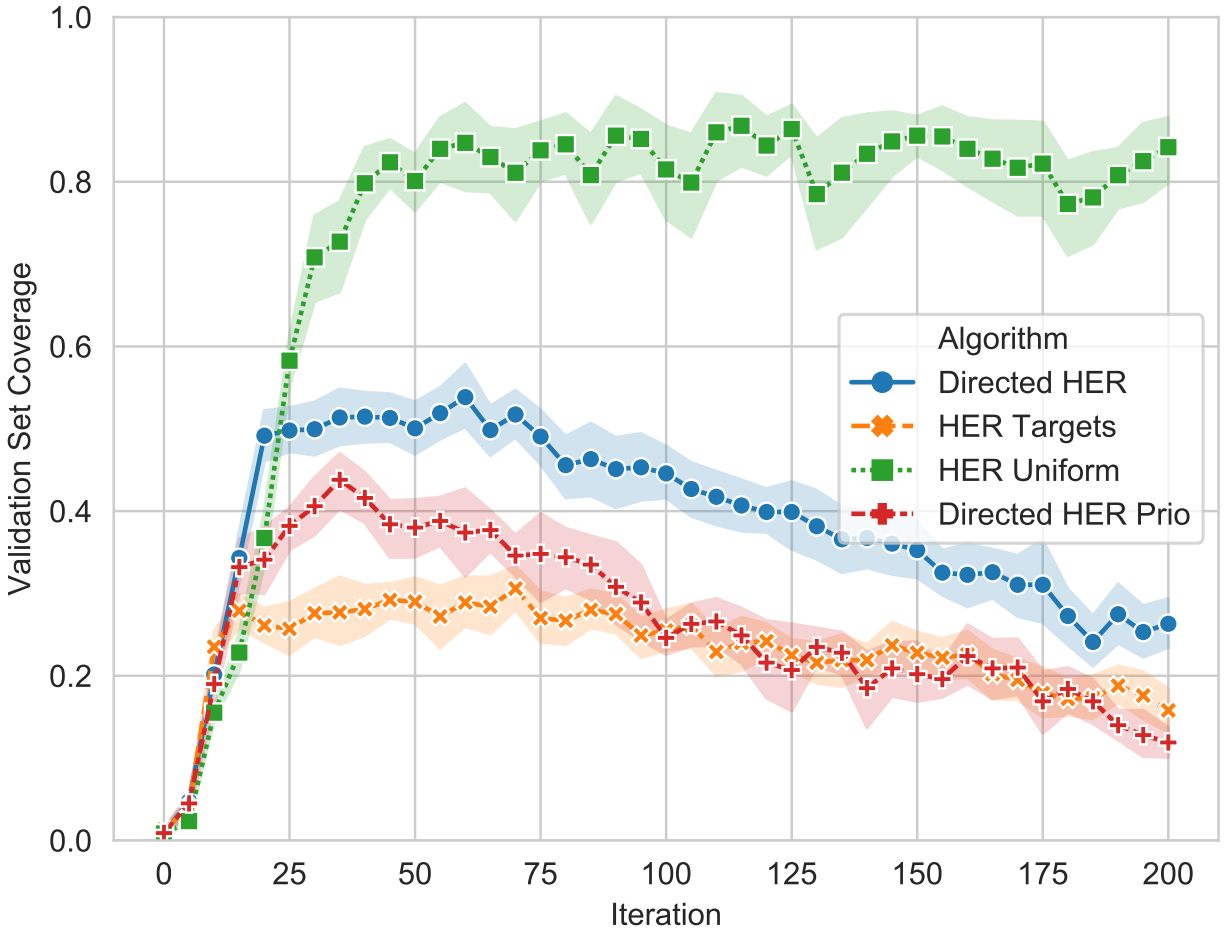


Figure 8: Coverage of the whole state space of the CrossGrid environment.

Experiment: Directed HER

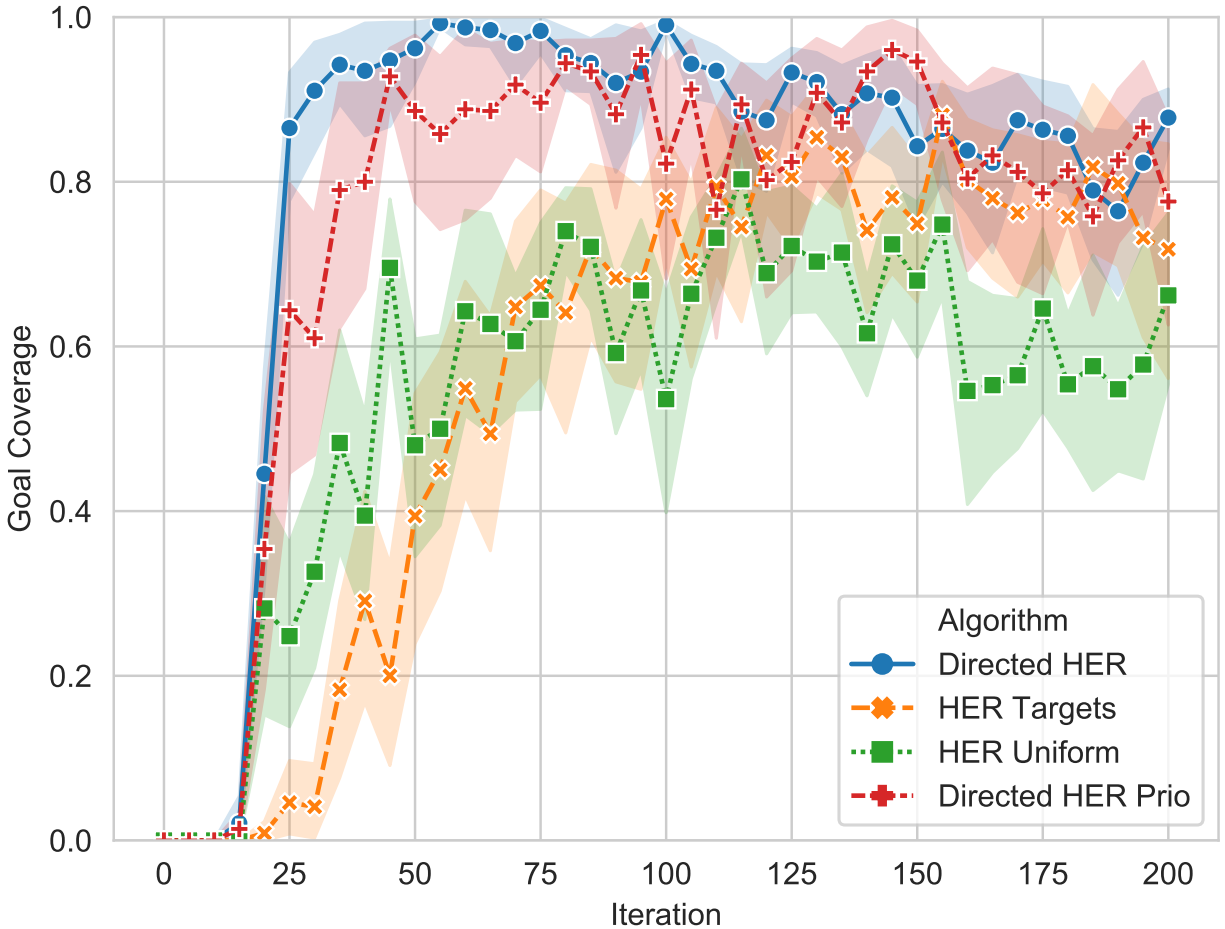


Figure 9: Coverage of the target region of the CrossGrid environment.

Experiment: Hand Movement "Thumbs-Up"

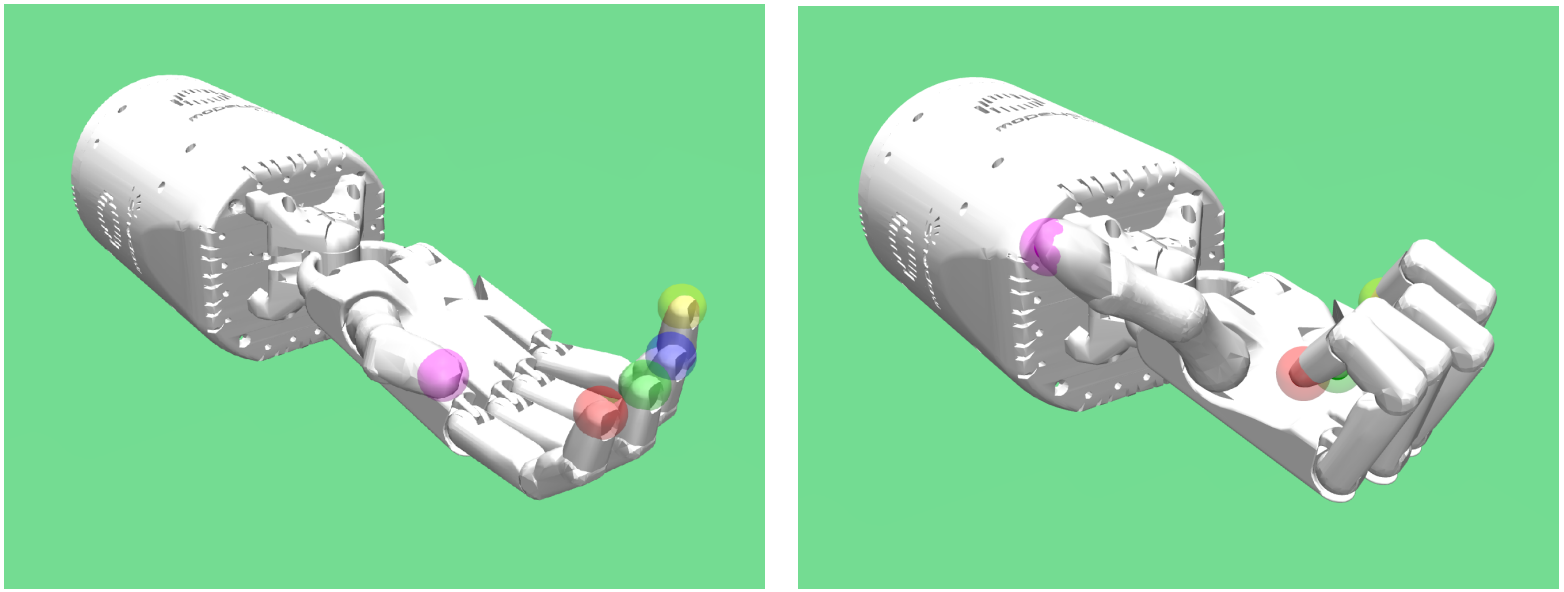
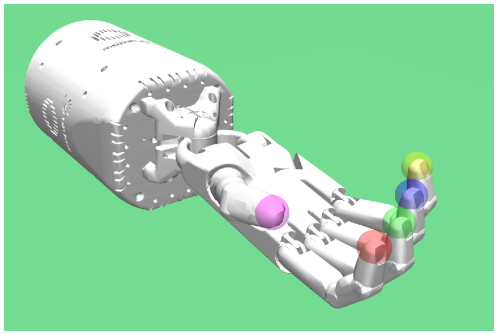
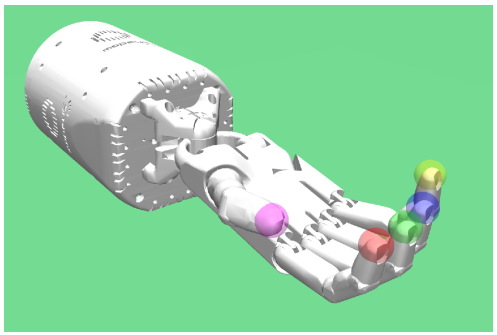
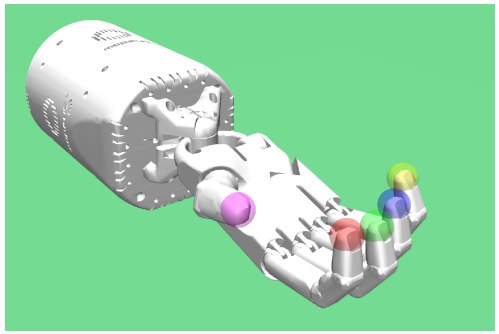


Figure 10: Left: Start Position, Right: Example of a target position

Experiment: Hand Movement "Thumbs-Up"

Does Directed HER generate a directed curriculum?

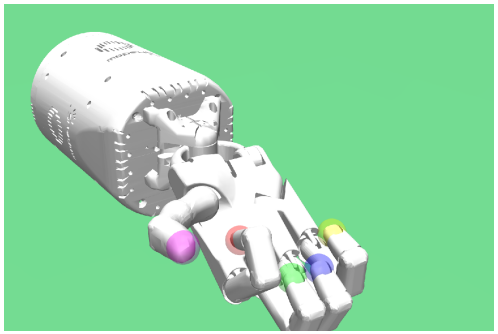
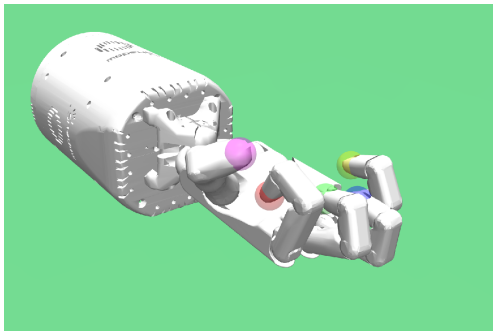
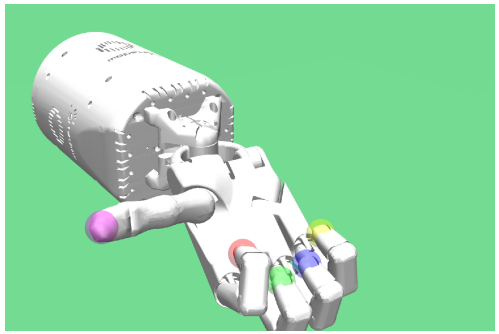
Epoch 1

Experiment: Hand Movement "Thumbs-Up"

Does Directed HER generate a directed curriculum?

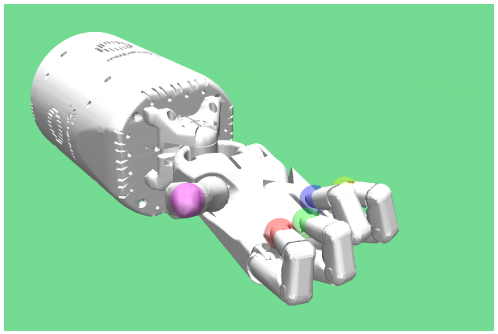
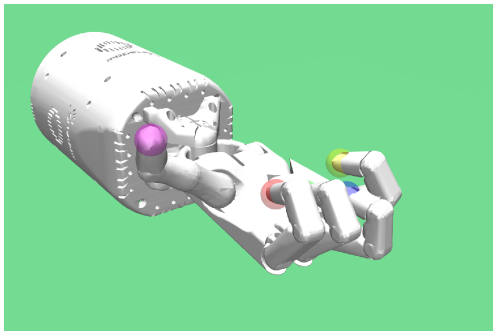
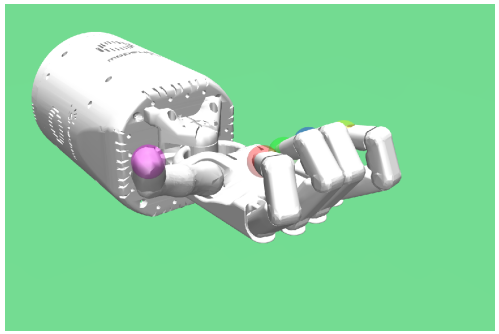
Epoch 5

Experiment: Hand Movement "Thumbs-Up"

Does Directed HER generate a directed curriculum?

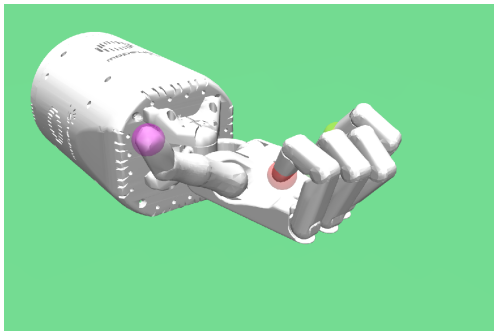
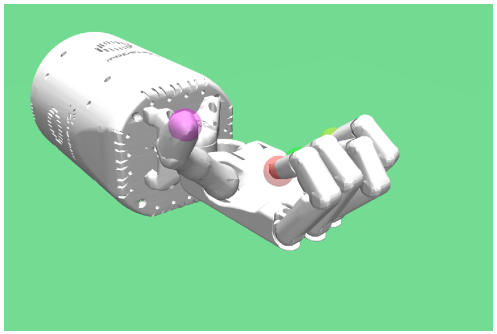
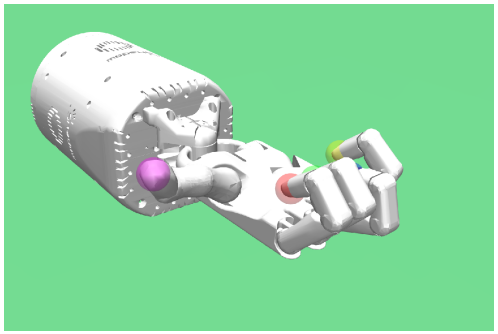
Epoch 9

Experiment: Hand Movement "Thumbs-Up"

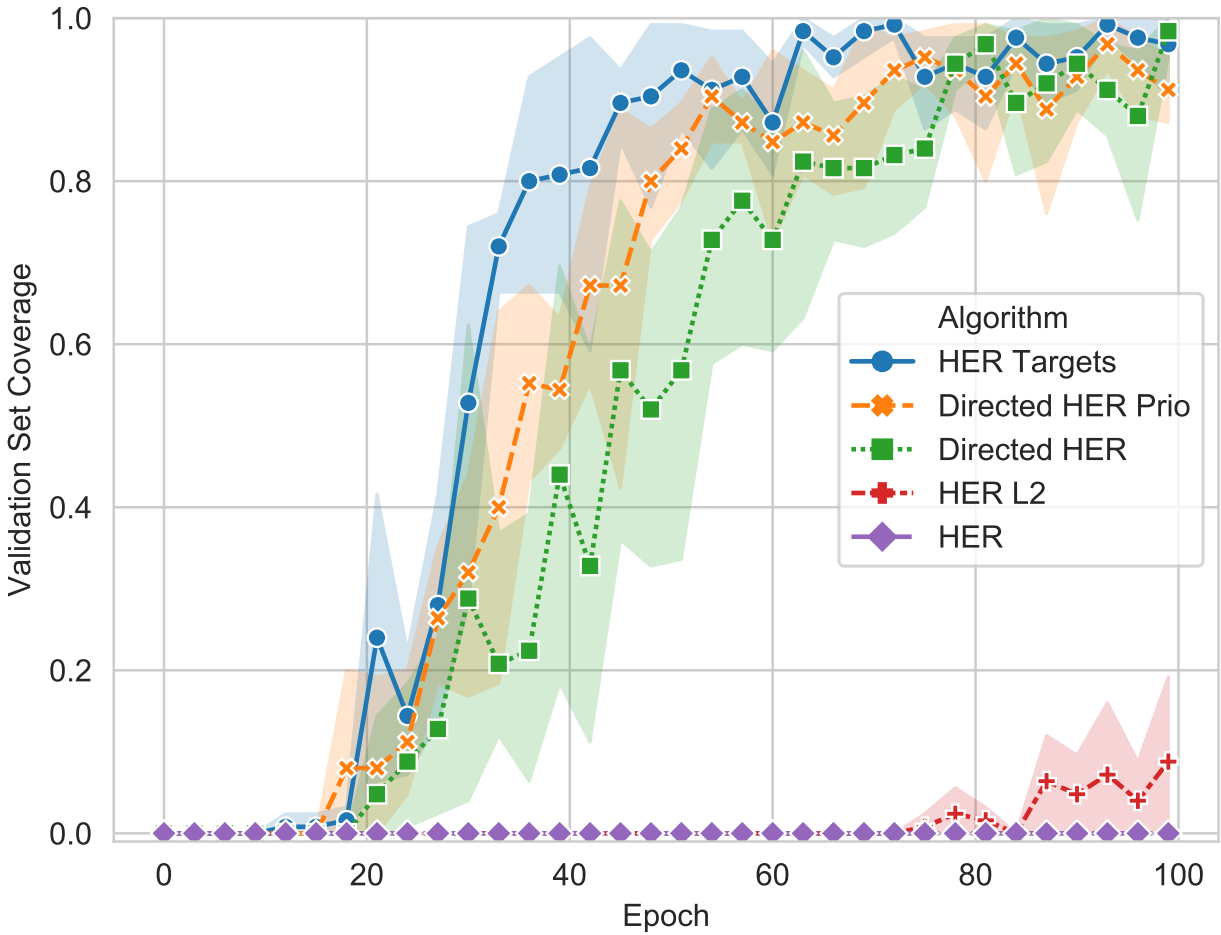
Does Directed HER generate a directed curriculum?

Epoch 11

Experiment: Hand Movement "Thumbs-Up"

Does Directed HER improve the performance of HER?



Experiment: Cube In-Hand Manipulation

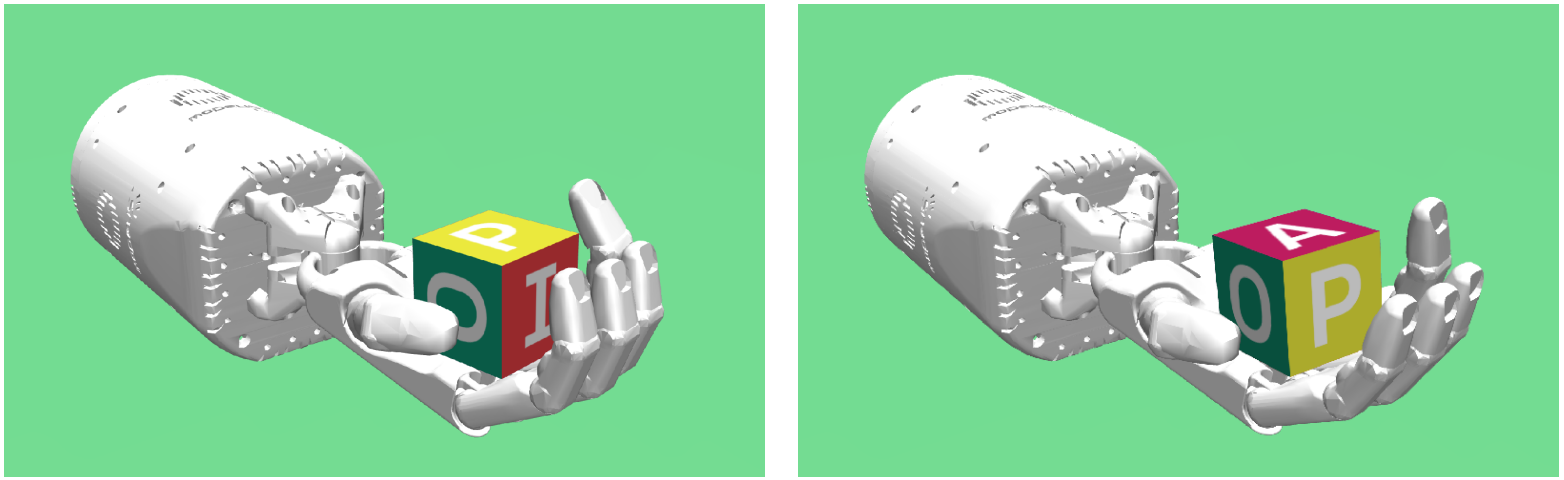
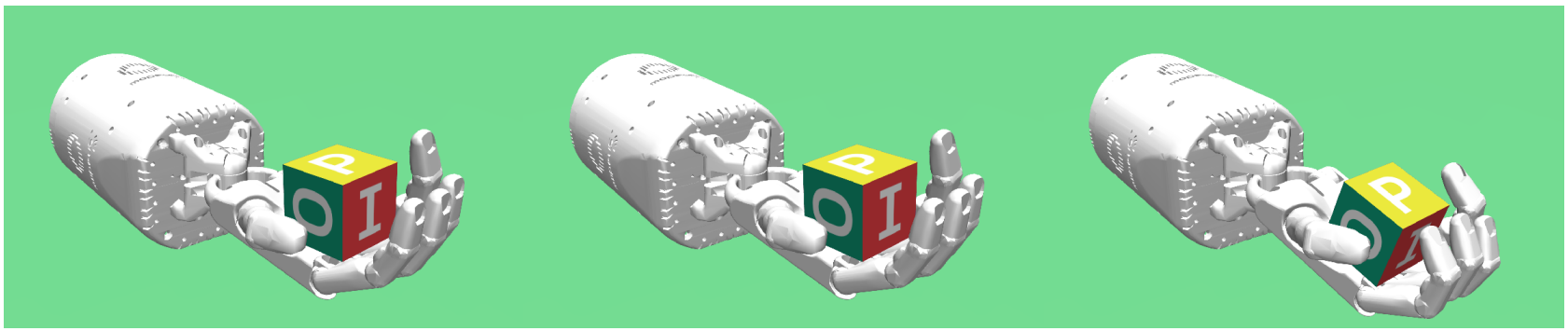


Figure 11: Left: Start Position, Right: Example of a target position

Experiment: Cube In-Hand Manipulation

Does Directed HER generate a directed curriculum?

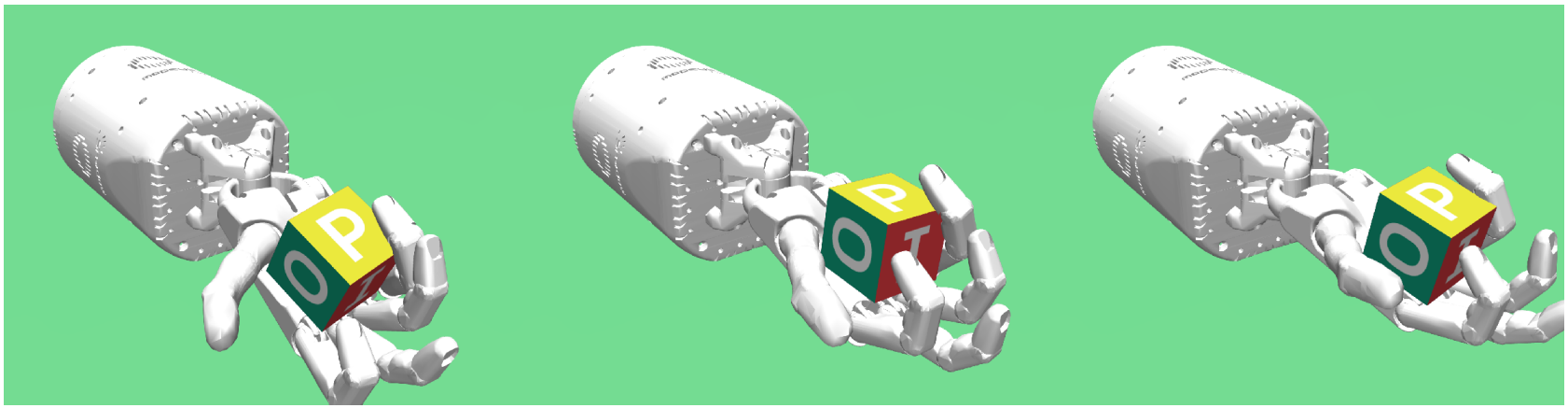
Epoch 1



Experiment: Cube In-Hand Manipulation

Does Directed HER generate a directed curriculum?

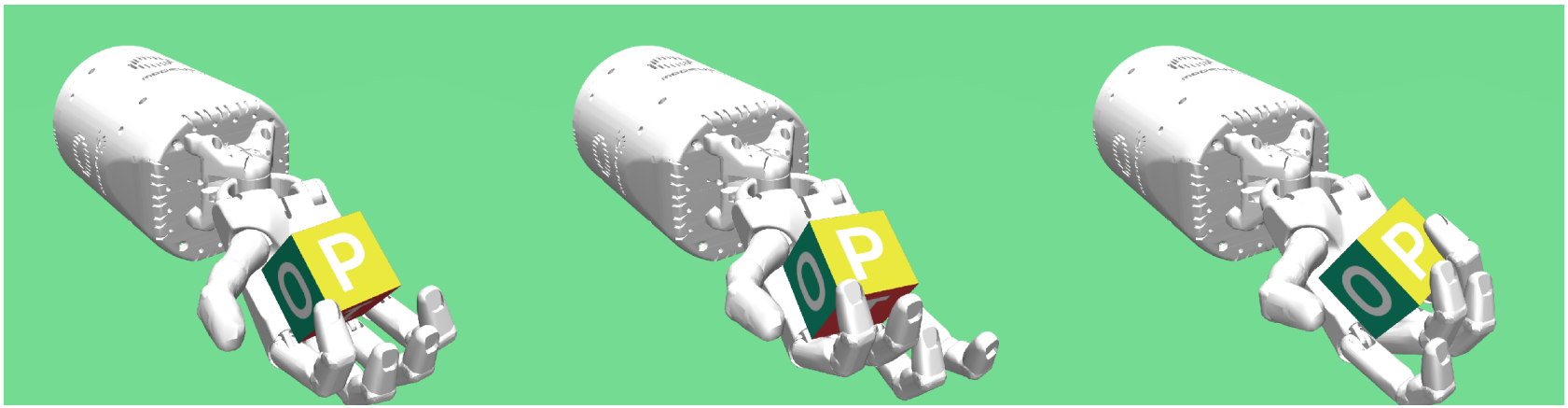
Epoch 21



Experiment: Cube In-Hand Manipulation

Does Directed HER generate a directed curriculum?

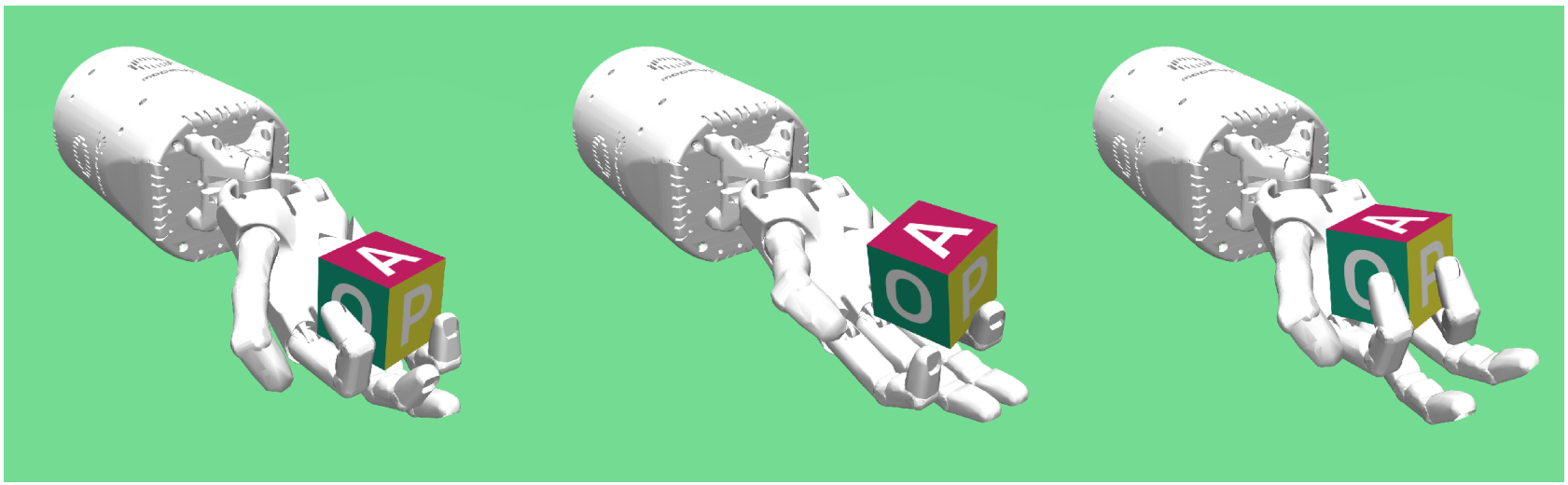
Epoch 41



Experiment: Cube In-Hand Manipulation

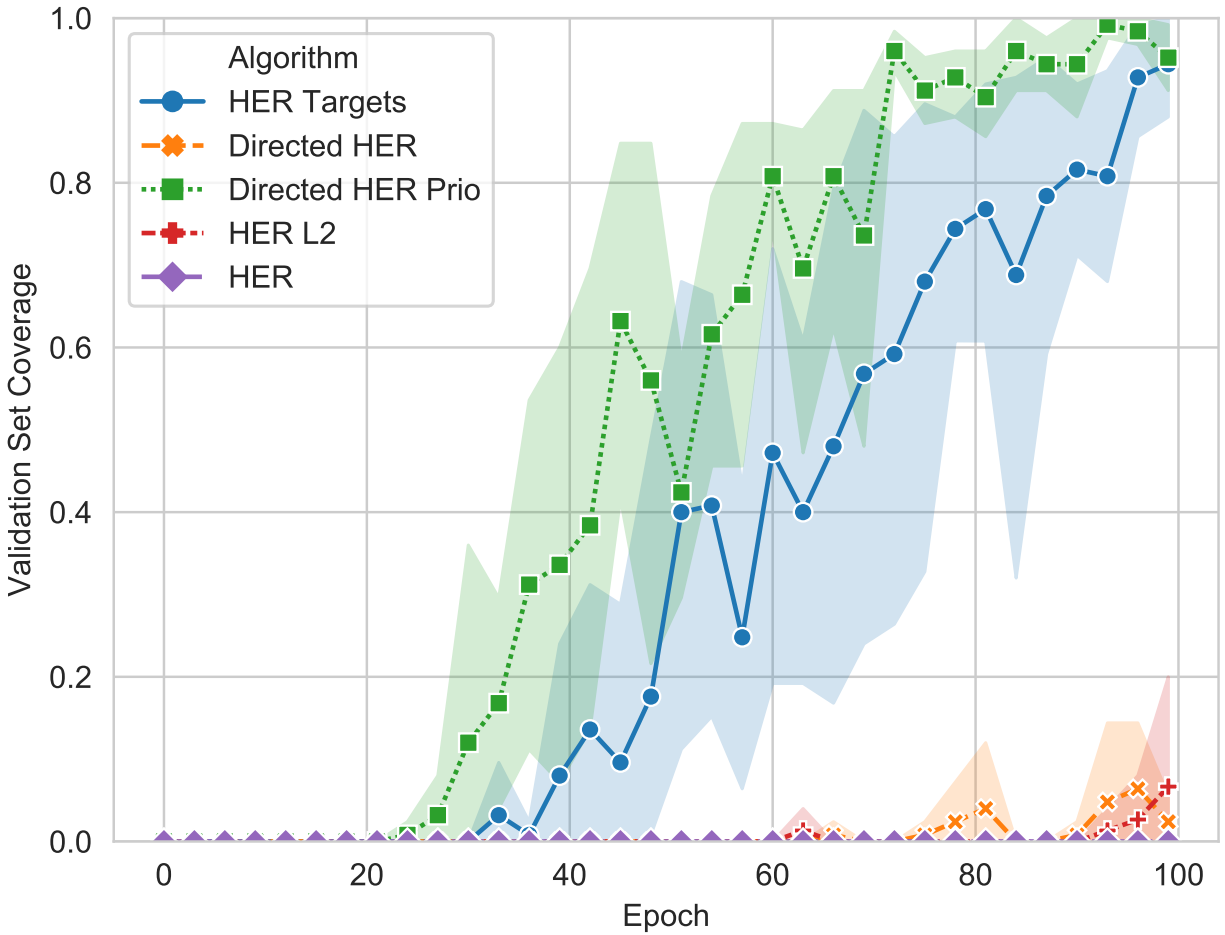
Does Directed HER generate a directed curriculum?

Epoch 51



Experiment: Cube In-Hand Manipulation

Does Directed HER improve the performance of HER?



Q & A

References

Andrychowicz, Marcin et al. (2017). "Hindsight experience replay". In: *Advances in Neural Information Processing Systems*, pp. 5048–5058.

Florensa, Carlos et al. (2017). "Reverse Curriculum Generation for Reinforcement Learning". In: *Conference on Robot Learning*, pp. 482–495.

Held, David et al. (2018). "Automatic goal generation for reinforcement learning agents". In: *International Conference on Machine Learning*.

Karpathy, Andrej and Michiel Van De Panne (2012). "Curriculum learning for motor skills". In: *Canadian Conference on Artificial Intelligence*. Springer, pp. 325–330.

Molchanov, Artem et al. (2018). "Region Growing Curriculum Generation for Reinforcement Learning". In: *arXiv preprint arXiv:1807.01425*.