EUROGRAPHICS
CONFERENCE 2024

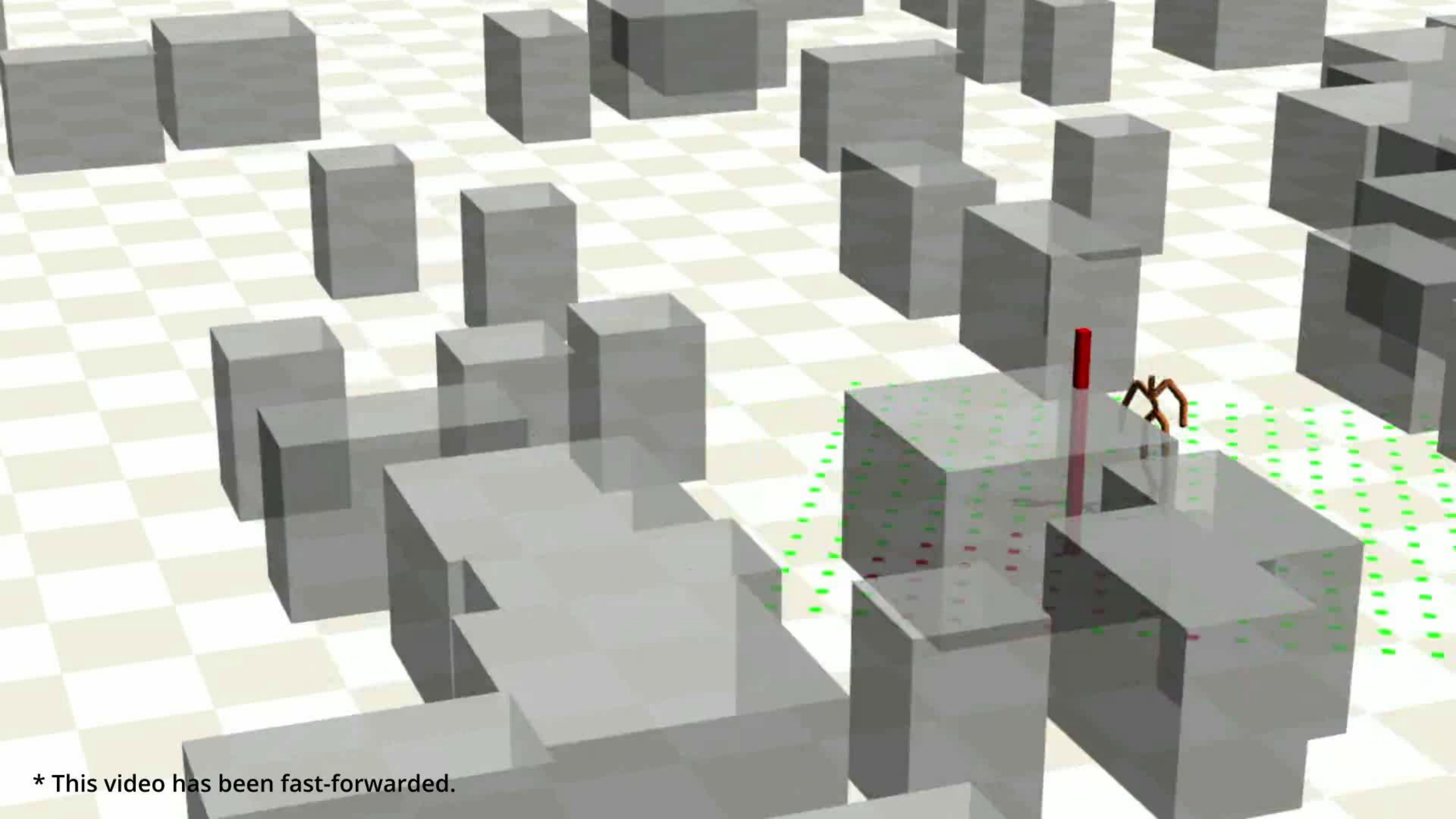
Utilizing Motion Matching with Deep Reinforcement Learning for Target Location Tasks

Jeongmin Lee

1st author

Jeongmin Lee[1,2], Taesoo Kwon[2], Hyunju Shin[1], Yoonsang Lee[2]

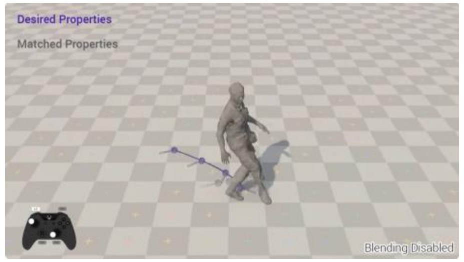
[1]Samsung Electronics Co. [2]Hanyang University


* This video has been fast-forwarded.

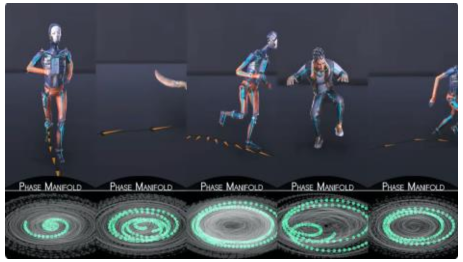
Related Work

- Motion matching
- Deep reinforcement learning

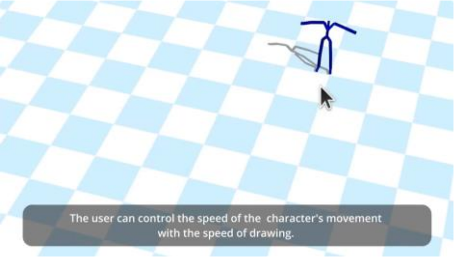
Learned
Motion Matching
[Holden 2017]

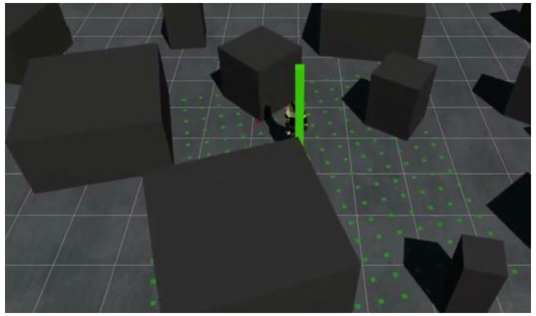


Deep Phase
[Starke 2022]

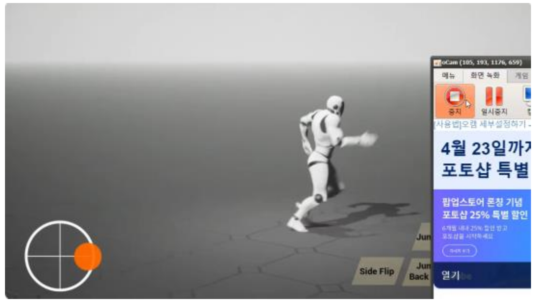


Interactive Character
Path-Following
Using Long-Horizon
Motion Matching
With Revised Future Queries
[Lee 2023]

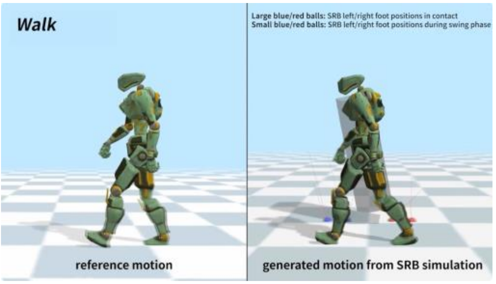




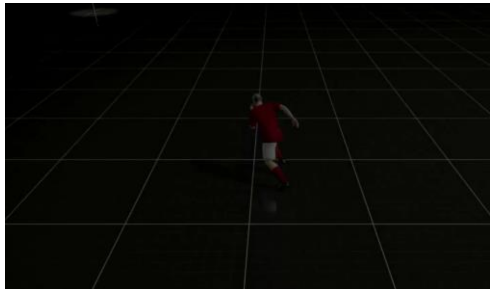
Motion Recommendation
For Online Character Control
[Cho 2021]



Learning Time-critical Response
for Interactive Character Control
[Lee 2021]



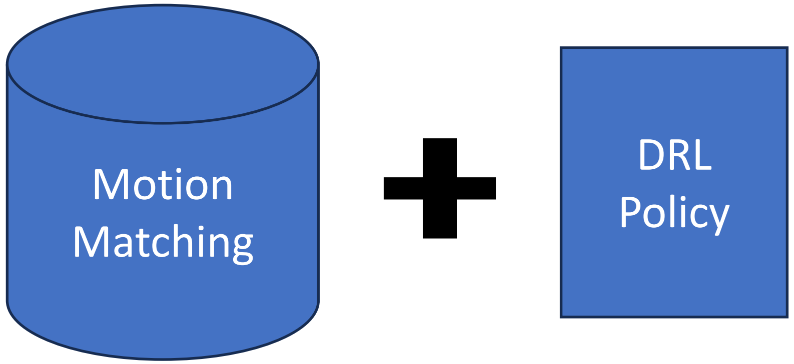
Adaptive Tracking
of a Single-Rigid-Body Character
in Various Environments
[Kwon 2023]



Motion VAE
[Ling 2020]

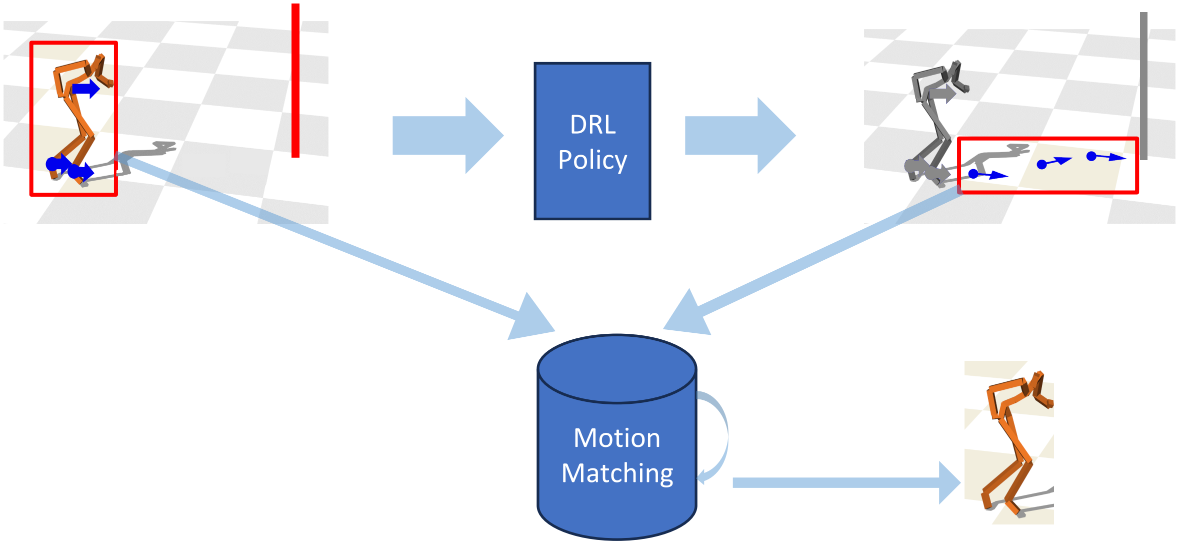
Motivation

- Train a policy using deep reinforcement learning(DRL) to generate character animation.
- Combine motion matching, with continuous state and action spaces.
- Enable direct generation of motion matching queries for long-term tasks, *target locations.*

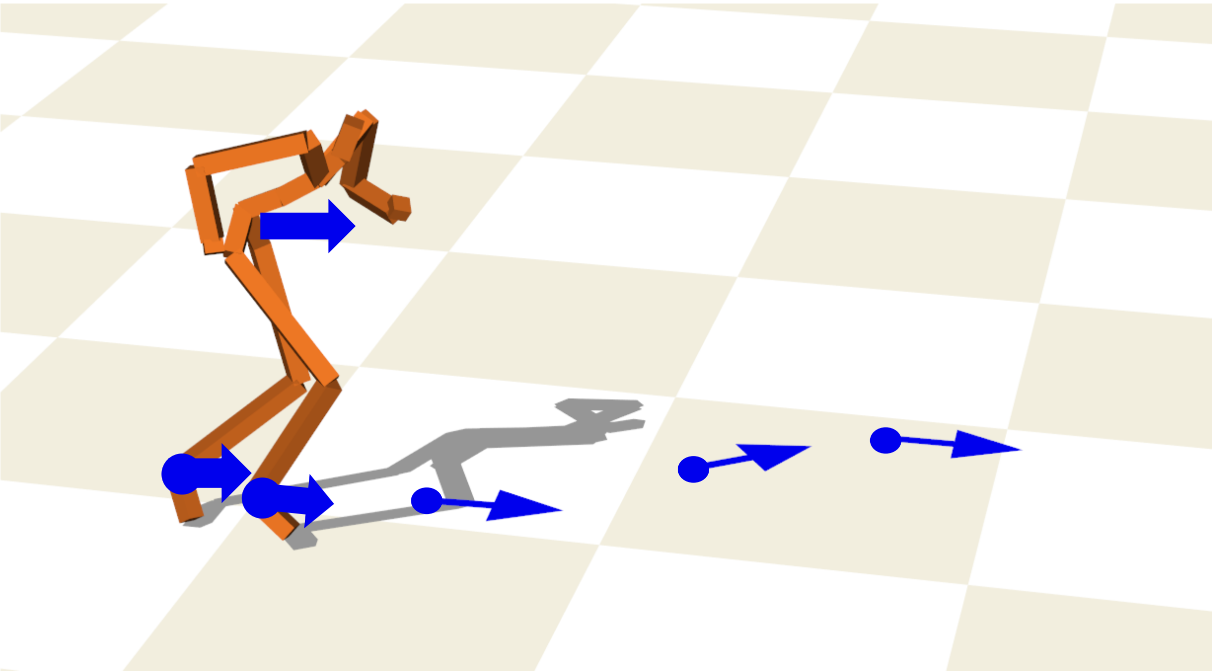


Our method

- Quickly learned within a short timeframe
- Available with a simple reward design
- A novel reward term and curriculum design to facilitate the learning (when with moving obstacles).

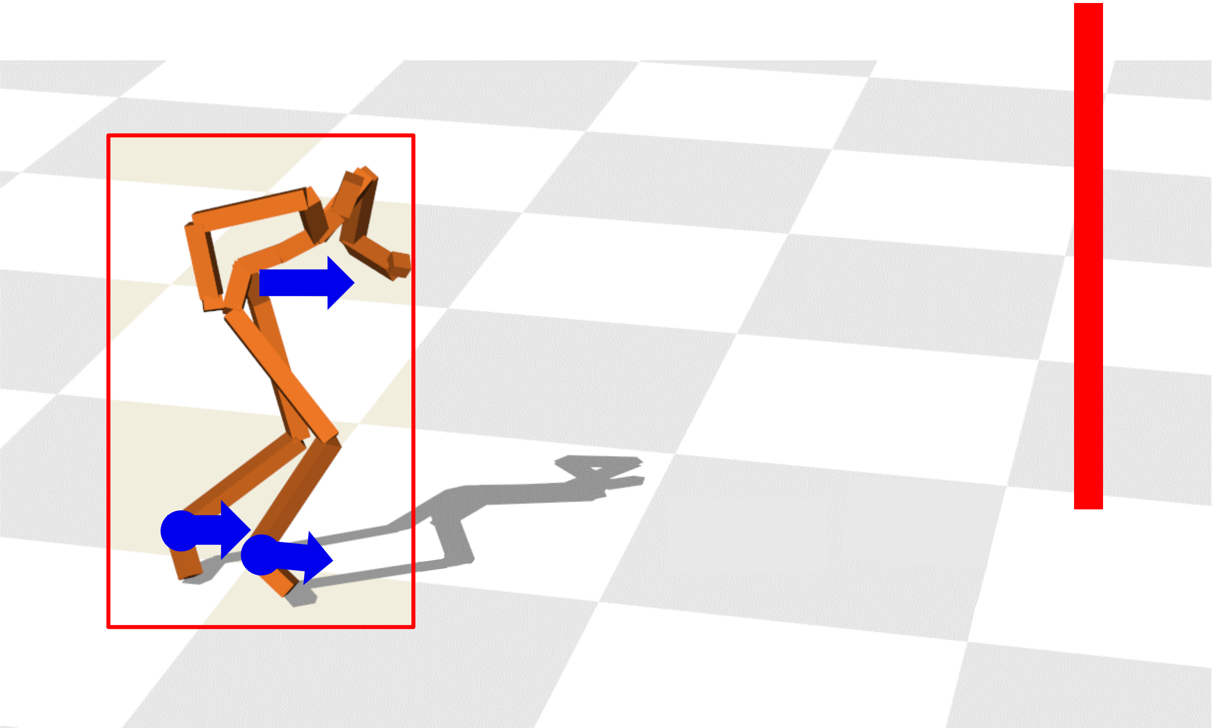


A quick summary of motion matching [GDC2016]



- Feature vector x at timestep t includes:
  - Character feature ($\boldsymbol{c}_{\mathrm{t}}$)
    - Foot positions
    - Foot velocities
    - Hip(root) velocity
  - Future trajectory ($\boldsymbol{t}_{\mathrm{t}}$)
    - Of the root node, 1-second length
- Search for the most similar posture frame j*.
  - $j^* = \operatorname{argmin}_j \left\| \hat{x} - x_j \right\|^2$

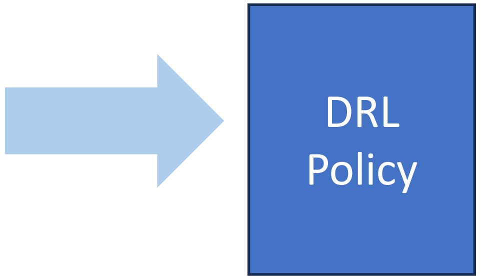
DRL Formulation with Motion Matching - State



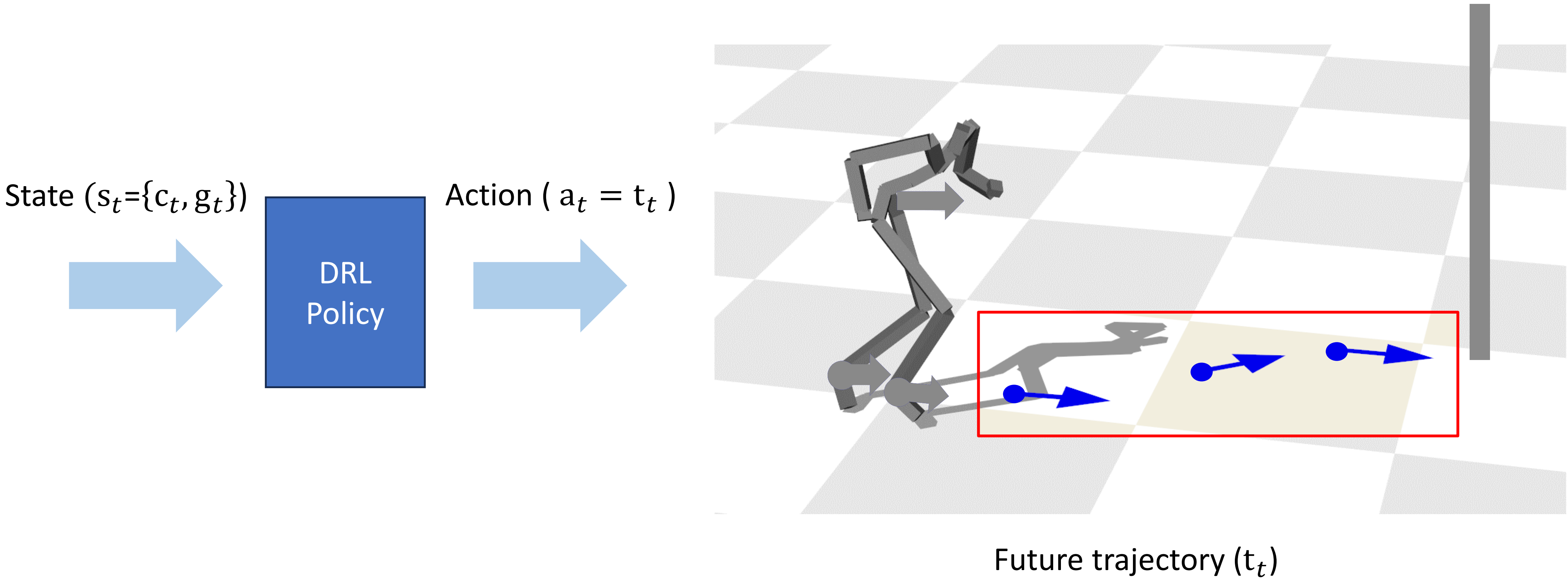
Character feature (c_t)

Goal (g_t)

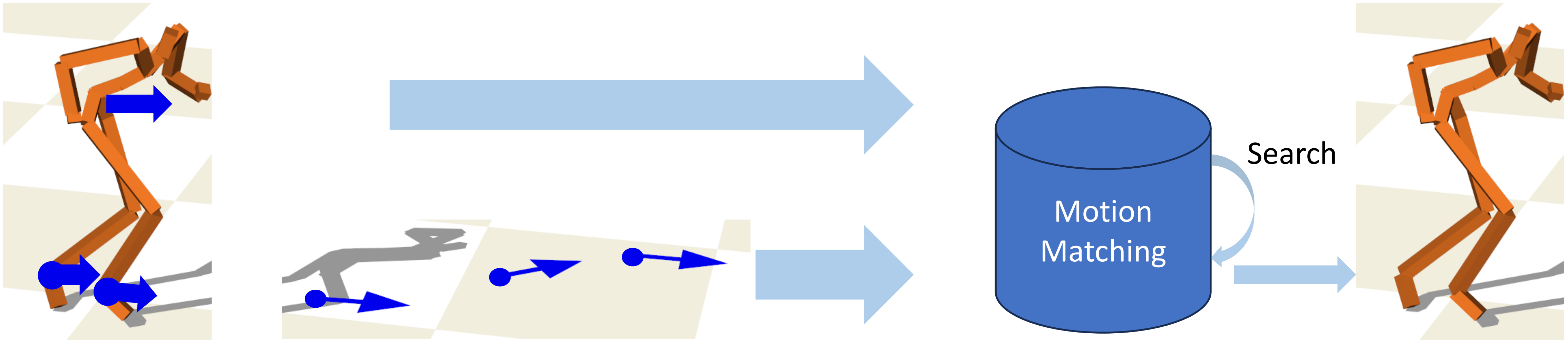
State (s_t=$\{c_t, g_t\}$)



DRL Formulation with Motion Matching - Action



DRL Formulation with Motion Matching – Motion Matching

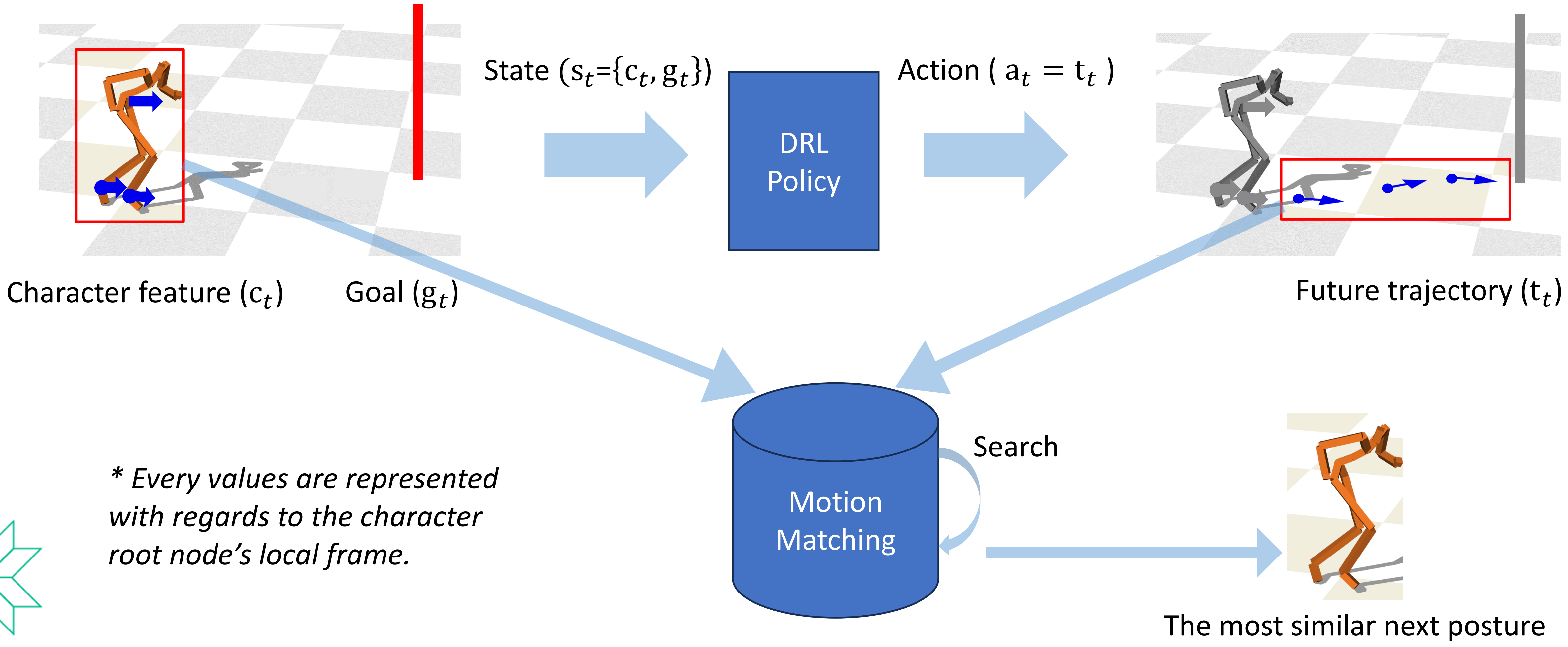


Character feature (c_t)

Future trajectory (t_t)

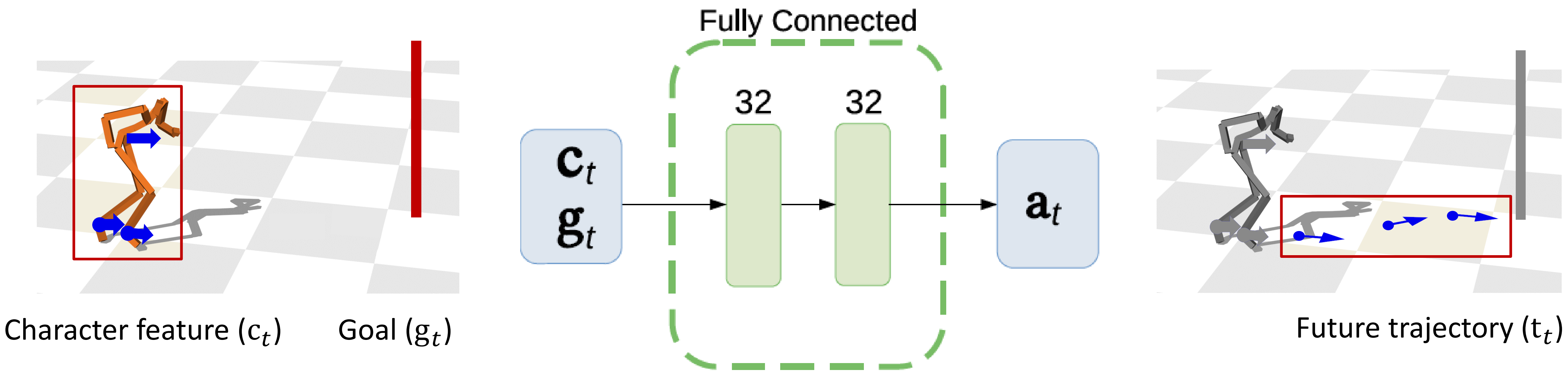
The most similar next posture

Overview of Our Method



** Every values are represented with regards to the character root node's local frame.*

Plane Environment

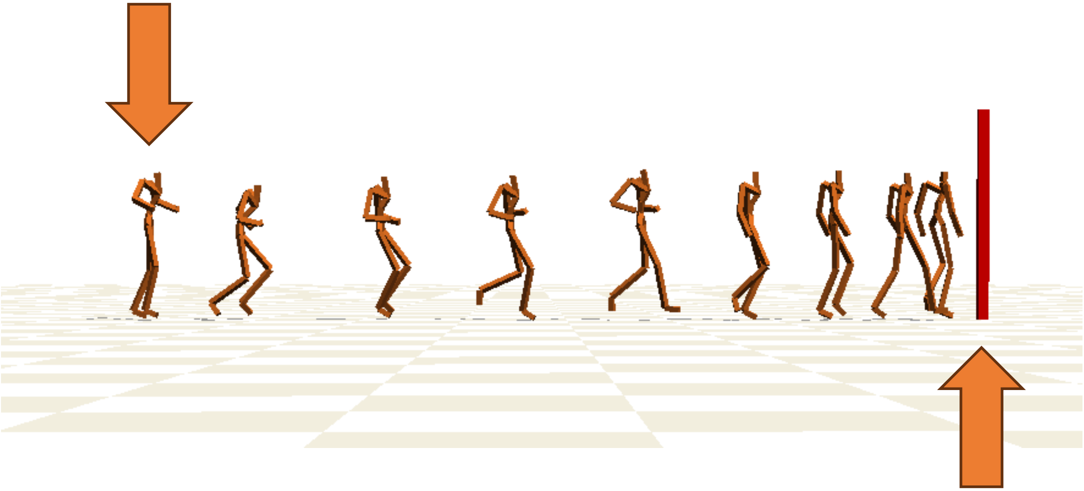


- Our plane environment has a bare plane with a sole target position.
- Our policy has two fully connected layers.
- At each training iteration, the policy is updated by rewarding a_t
- Reward $r_t = \exp(-\|g_t\|_2)$, the character root's distance to the goal position.

Training for Plane Environment

- Facilitate learning process by..
  - Initializing to a random posture frame.
  - Giving extra bonus rewards when an episode is "*successful*", and we terminate the episode right away.

Initializes to a random frame.



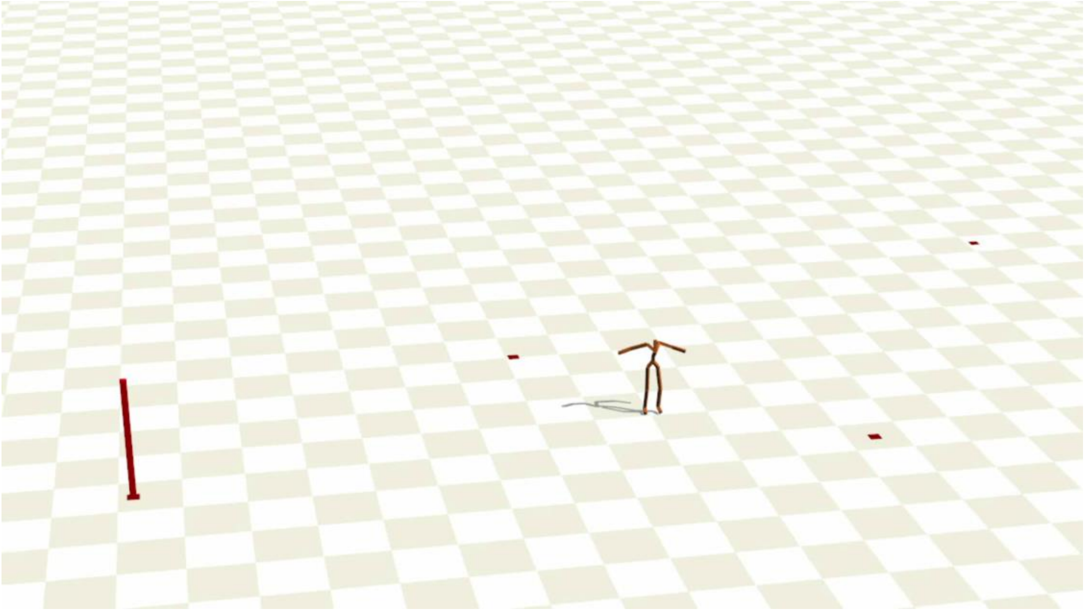
Terminates when,

1. The goal is close enough ($\|g_t\|_2 < 0.5$)
2. Or the episode is long enough ($t >= 200$)

Experimental Results – Plane Environment

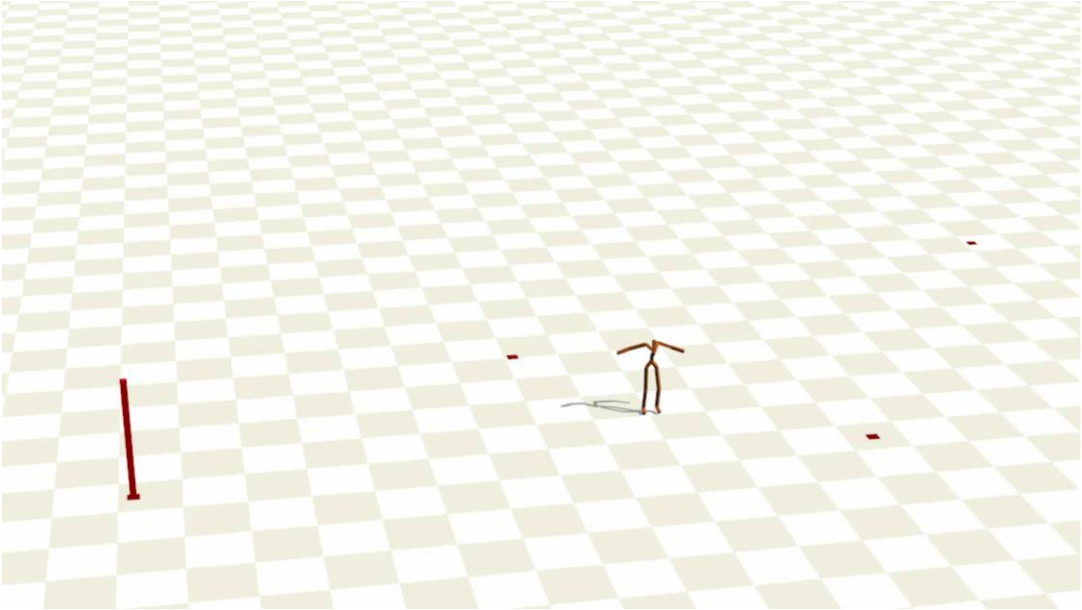
After 30 seconds of training (20K steps)

- Slow, less accurate.
- Head towards the target.



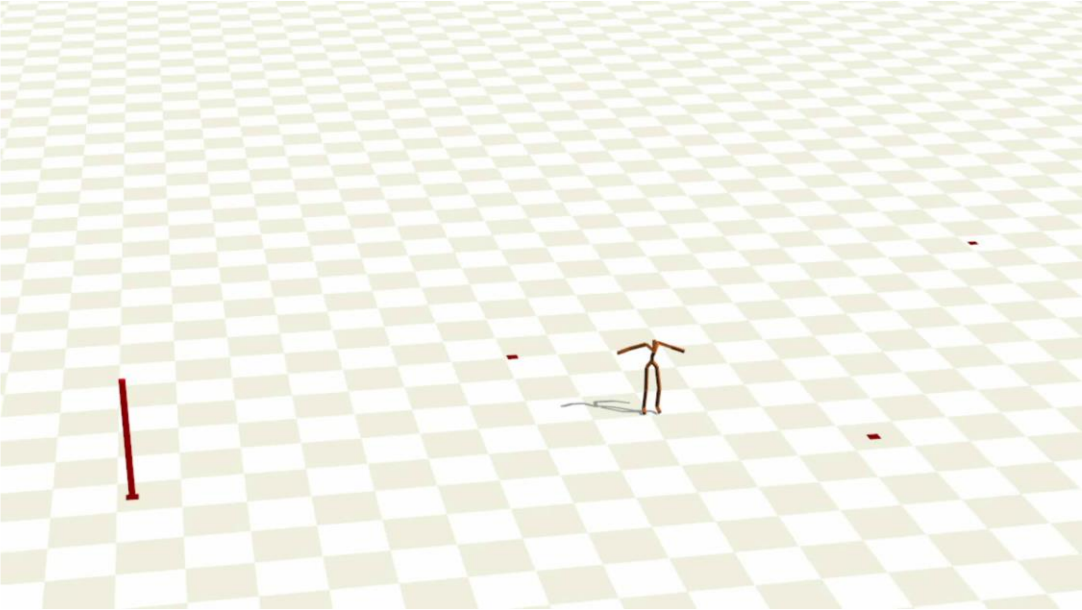
After 150 seconds of training (100K steps)

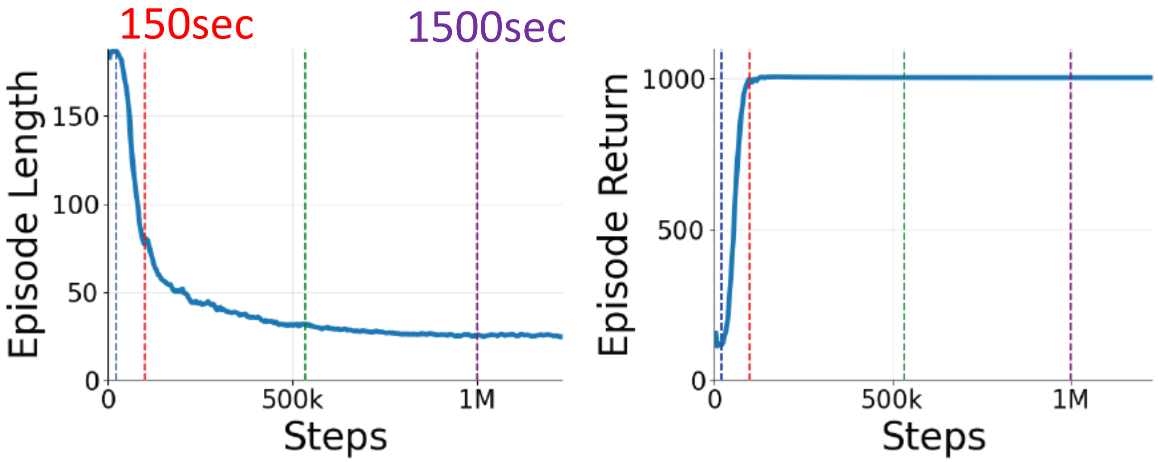
- Every episodes are successful.



Experimental Results – Plane Environment

- After then, the policy learns how to reach the target faster.
- After 1500 seconds of training (1M steps)



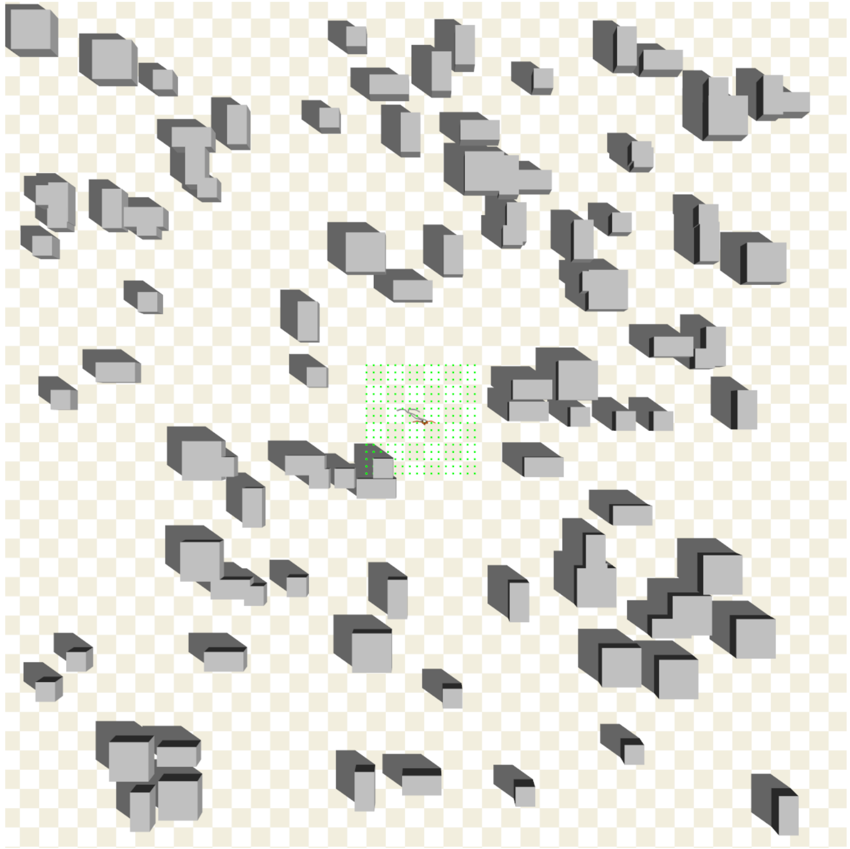


(a) (b)

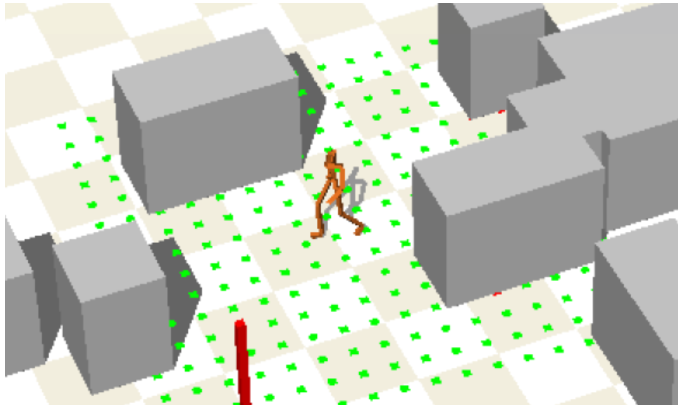
Extensions for Moving Obstacles Environment

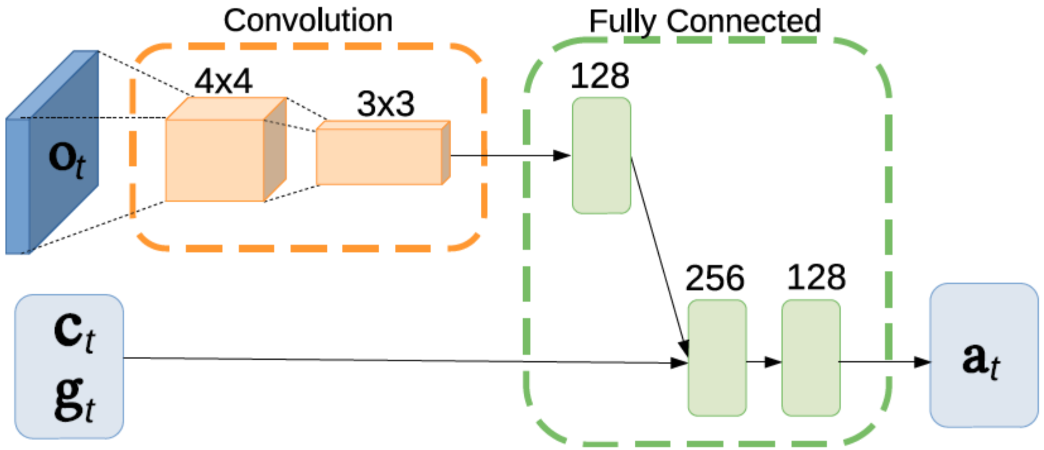
- Obstacles
  - 100 random obstacles

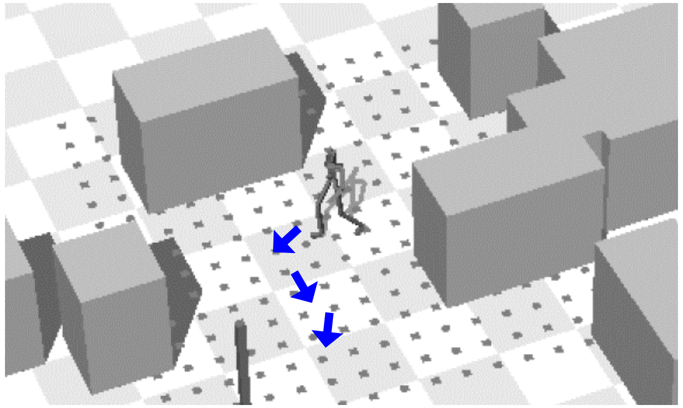
- Episode terminates when,
  1. Character-obstacle collision
  2. Or the episode is long enough
     - $t > 1000$



Extensions for Moving Obstacles Environment



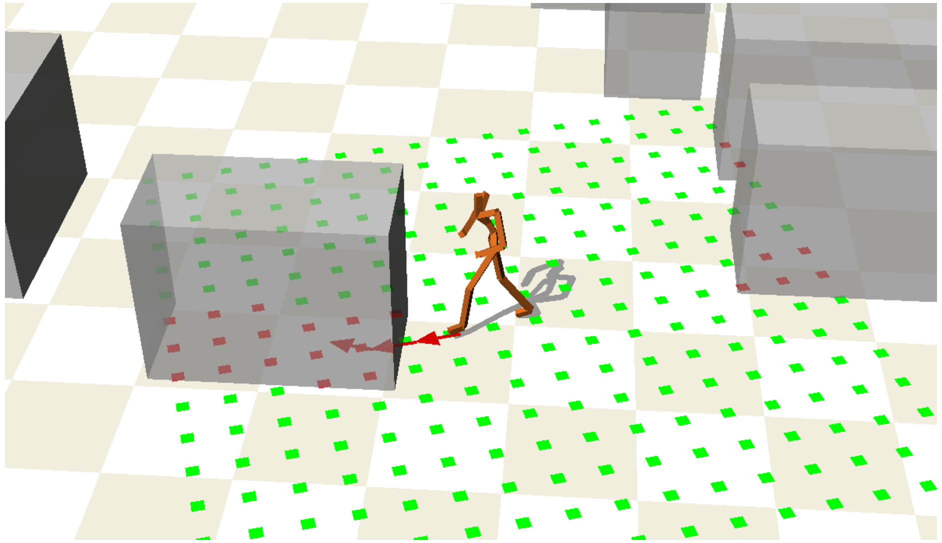




- 2 dimensional sensor $\boldsymbol{o}_t$ is given as an additional observation.
- The output action is the future trajectory $\boldsymbol{t}_t$, the same as the Plane Environment.

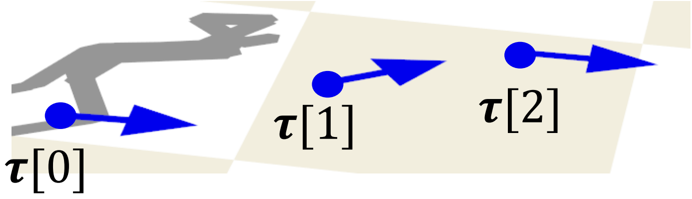
Extensions for Moving Obstacles Environment
(1) Hit Reward

- A novel reward term
  - A penalty on the count of future trajectory positions in an action that intersect with any obstacle.



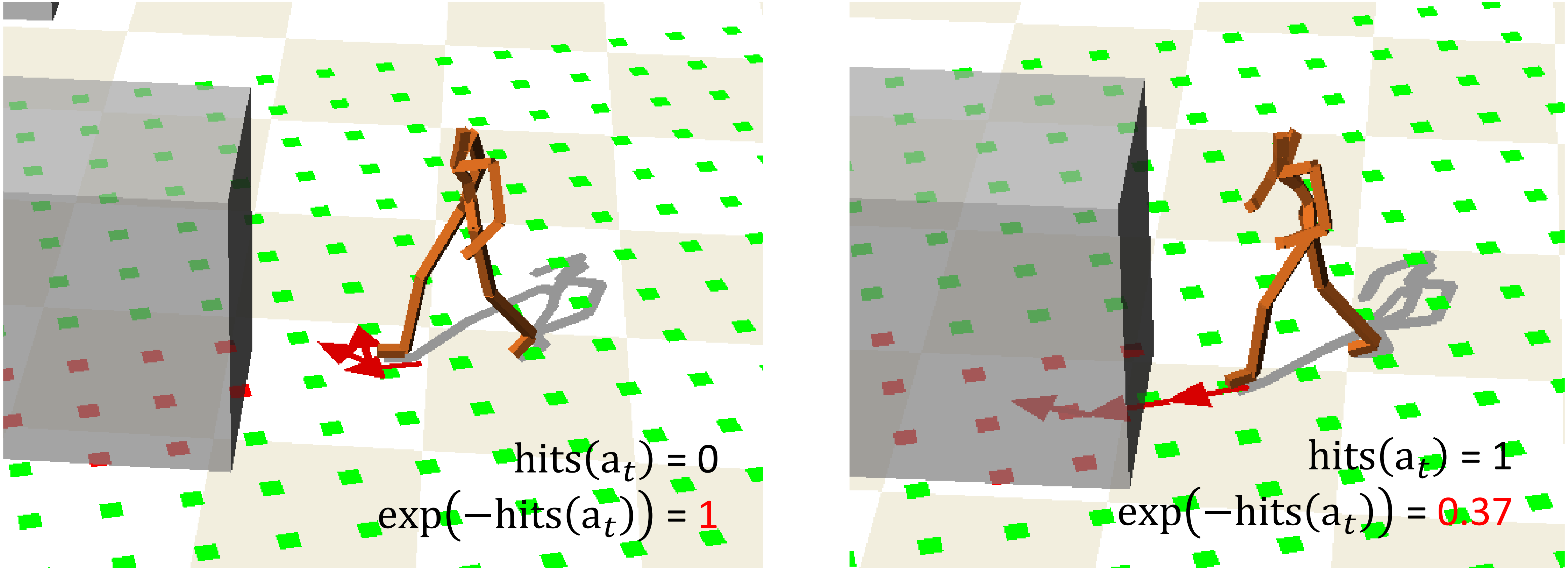
$$r_t = \exp(-\|g_t\|_2) + \exp(-\mathrm{hits}(a_t))$$

$$\mathrm{hits}(\mathbf{a}_t) = \sum_{k=0}^{2} \begin{cases} 1 & \text{if } \boldsymbol{\tau}[k] \text{ is inside any obstacle} \\ 0 & \text{else.} \end{cases}$$



* Future trajectory positions $\boldsymbol{\tau}$

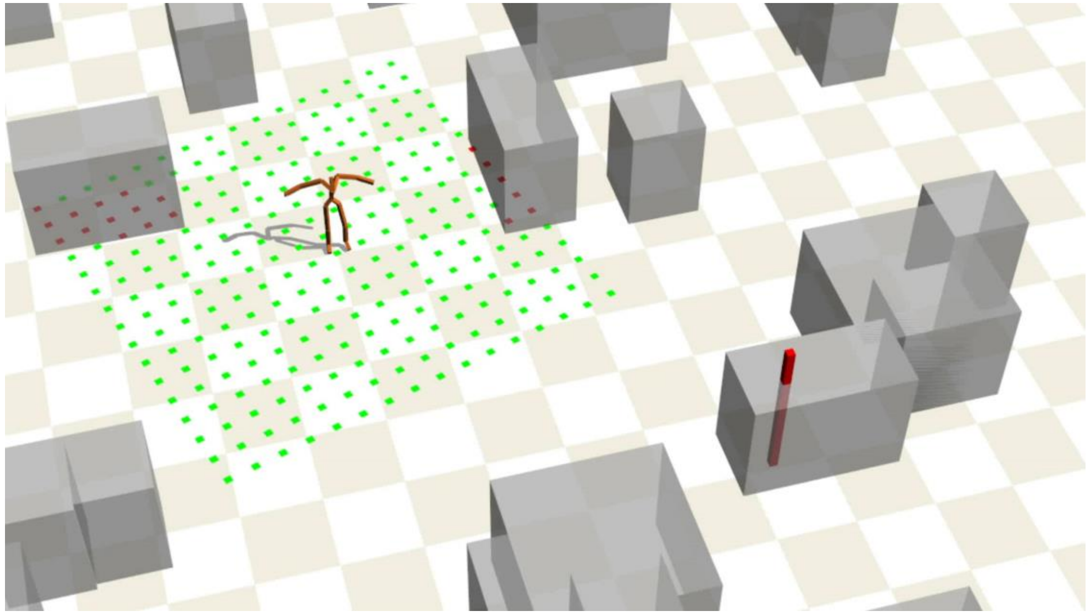
Extensions for Moving Obstacles Environment
(1) Hit Reward



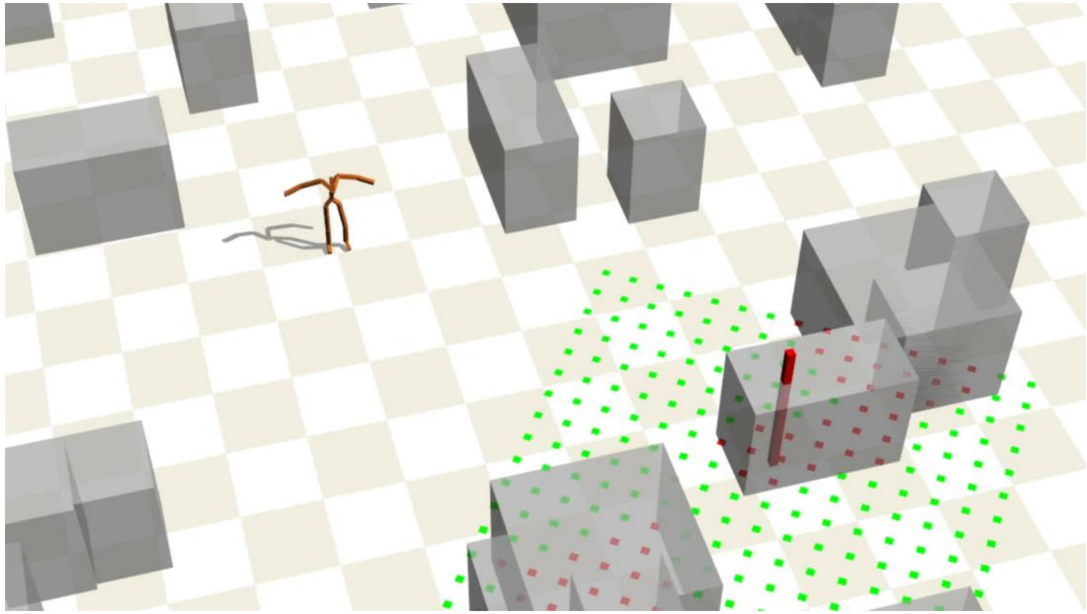
The action with future trajectory position intersecting with an obstacle gets less rewards.

Extensions for Moving Obstacles Environment
(1) Hit Reward

Policy with *hit reward*



Policy without *hit reward*

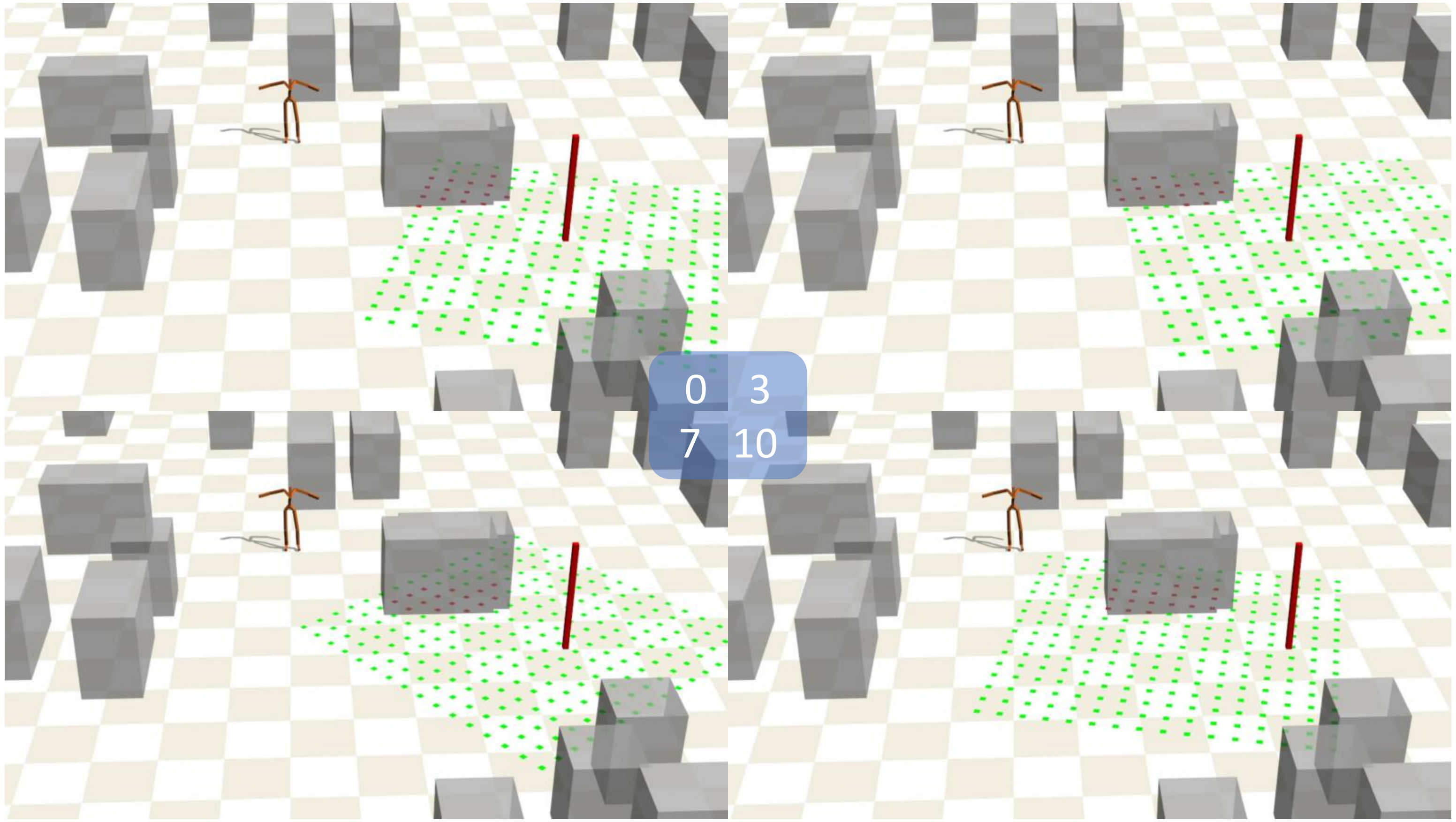


Extensions for Moving Obstacles Environment
(2) Obstacle Curriculum

- Enhances the learning of policies.
- Gradually increases
  - Goal sample area
  - Obstacle speeds
- Progress to the next level whenever
  - Over 40% of episodes collected in this iteration are "*successful*".

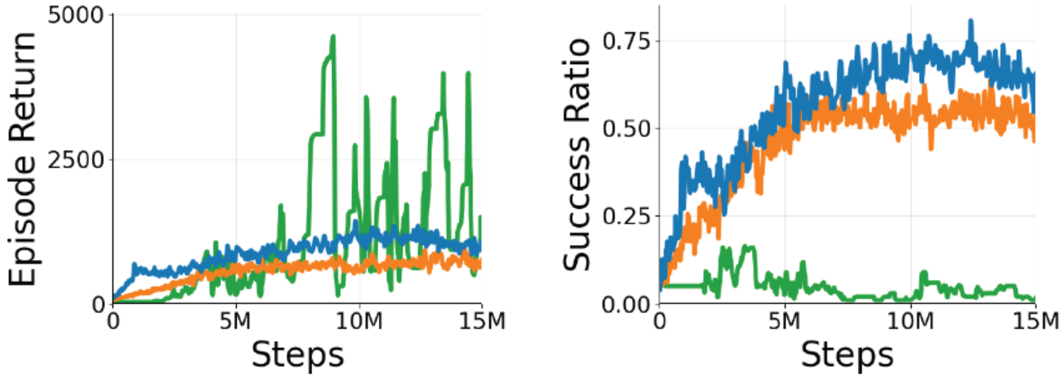
	Obstacle speed	Sample area of g_t
Level0	x 0	5m x 5m
Level1	x 0.1	5.5m x 5.5m
Level2	x 0.2	6m x 6m
Level3	x 0.3	6.5m x 6.5m
Level4	x 0.4	7m x 7m
Level5	x 0.5	7.5m x 7.5m
Level6	x 0.6	8m x 8m
Level7	x 0.7	8.5m x 8.5m
Level8	x 0.8	9m x 9m
Level9	x 0.9	9.5m x 9.5m
Level10	x 1.0	10m x 10m

Extensions for Moving Obstacles Environment
(2) Obstacle Curriculum



- Policy in environments of different levels.
- The policy is exposed to more difficult task levels while the learning proceeds.

Experimental Results – Moving Obstacles Environment



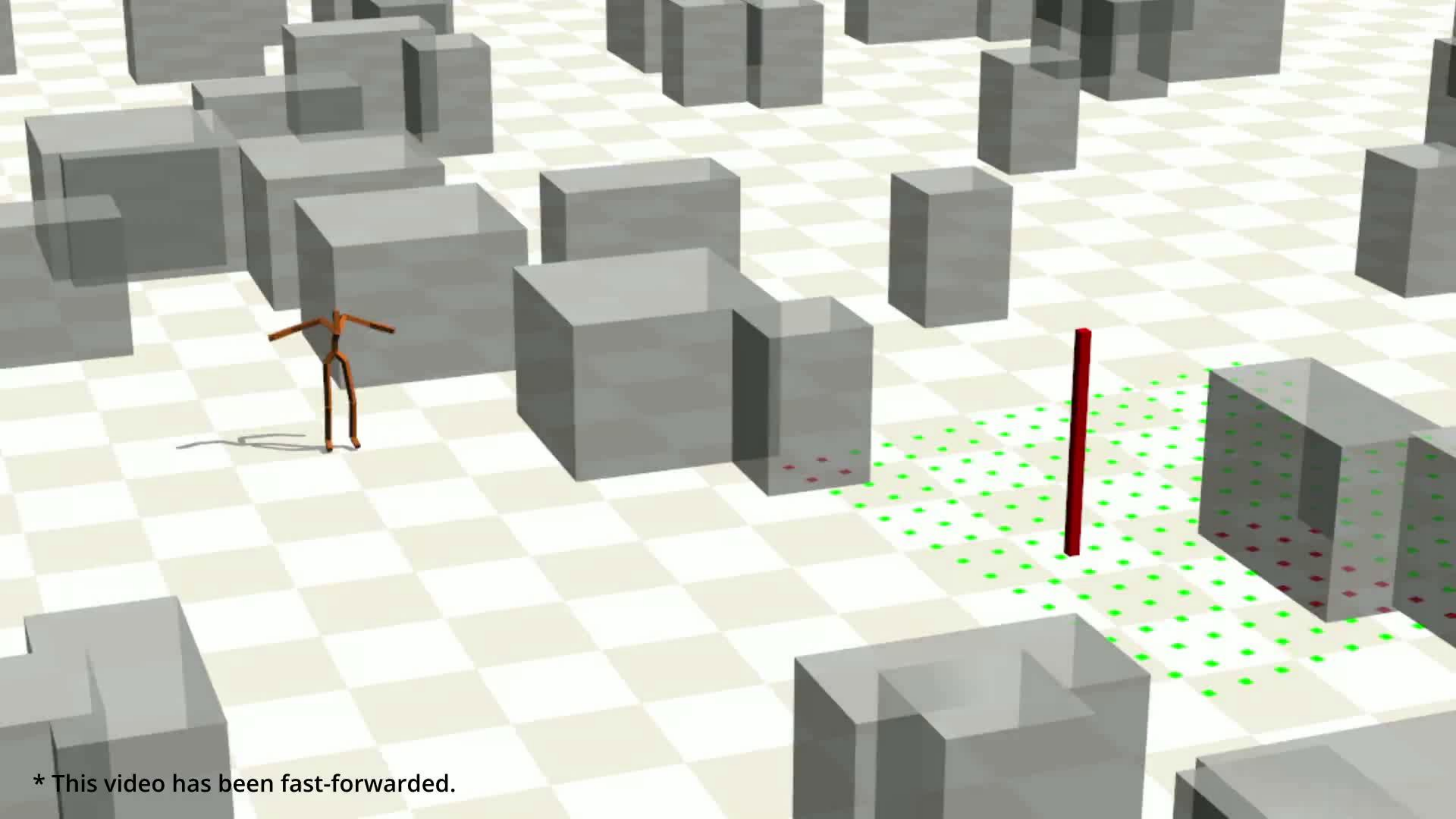
(a) (b)

- Ours
- Without hit reward
- Without obstacle curriculum

- Ours converges to the highest success ratio.
- Hit reward makes our method stable.
- Obstacle curriculum facilitates the learning.

* Episode Return: Mean of episode returns per iteration.

* Success Ratio: The rate of episodes collected in this iteration that successfully reached the target location.



* This video has been fast-forwarded.

Thank you.

More Information:

Supplementary video: https://youtu.be/kR47MrPhJGk

https://cgrhyu.github.io/publications/2024-matching-drl.html

Contact Information:

j0064423@gmail.com