Homework #2
EE 641: Fall 2024

Name: Nissanth Neelakandan Abirami
USC ID: 2249203582
Instructor: Dr. Franzke

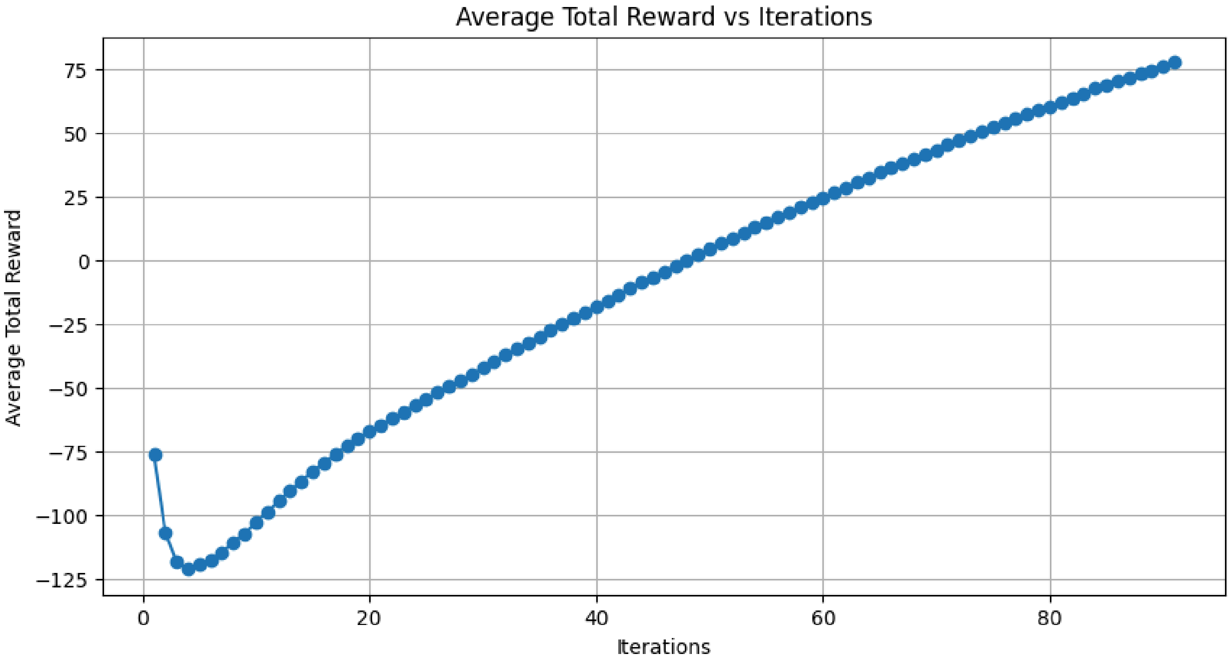
Due to GPU Constraints and Time , I couldn't explore multiple combination if hyperparameters but tried my level to include most cases

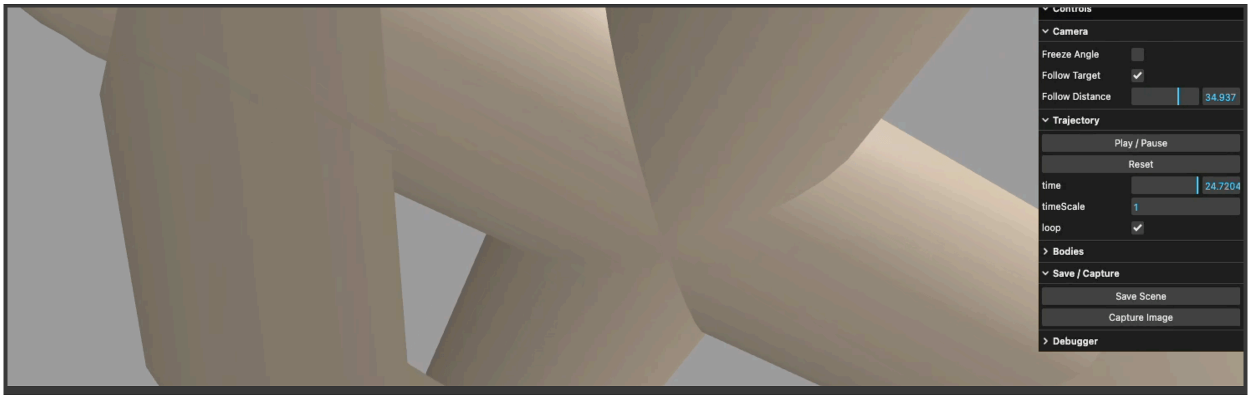
1)

ANT

Hyperparameter 1 :

Learning Rate : 3e-4
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.80
Gae Lambda : 0.95
Update Epochs : 10
Clip Coef : 0.4
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.0
Vf Coef : 0.5
Max Grad Norm : 0.5
Target Kl : None
Num Mini Batches : 32
Total Timesteps : 3000000
Minibatch Size : 200
Batch Size : 400
Num Iterations : 100

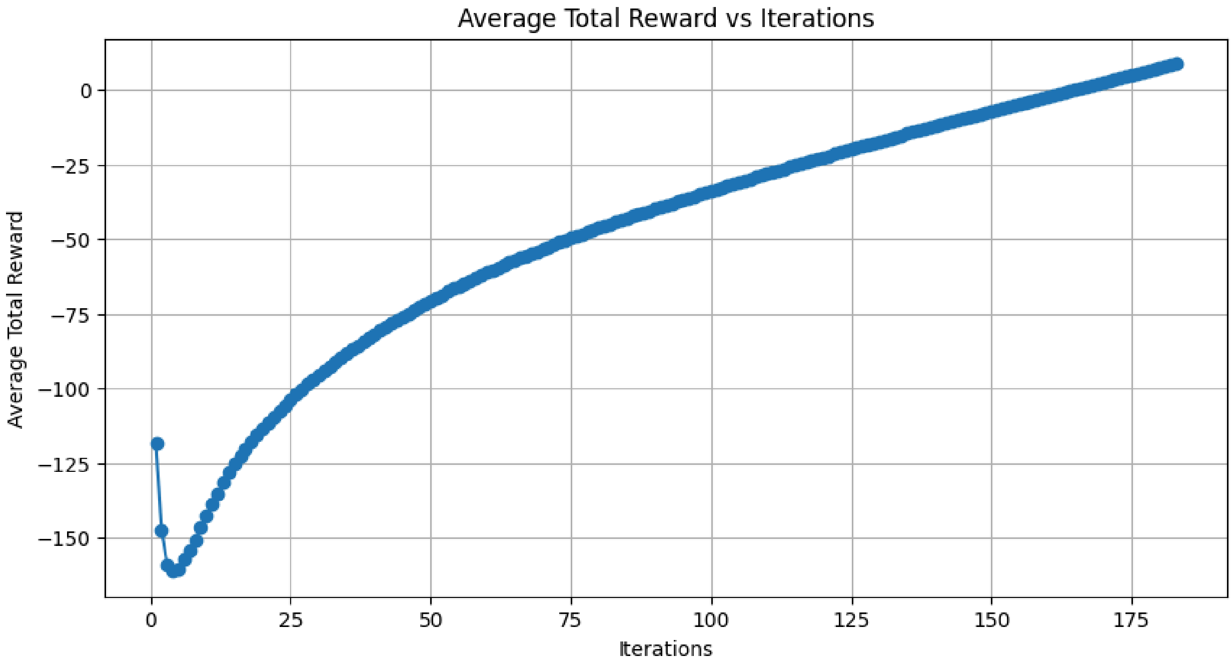


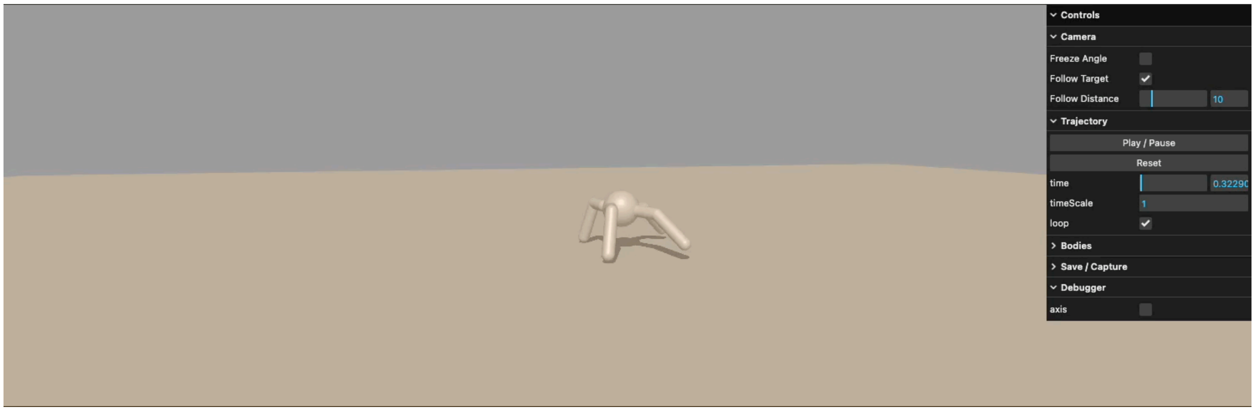


Freeze Angle □
Follow Target ☑
Follow Distance 34.937
time 24.7204
timeScale 1
loop ☑

Hyperparameter 2 :

Learning Rate : 3e-5
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.78
Gae Lambda : 0.95
Update Epochs : 10
Clip Coef : 0.37
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.0
Vf Coef : 0.5
Max Grad Norm : 0.5
Target Kl : None
Num Mini Batches : 32
Total Timesteps : 6000000
Minibatch Size : 100
Batch Size : 200
Num Iterations : 200

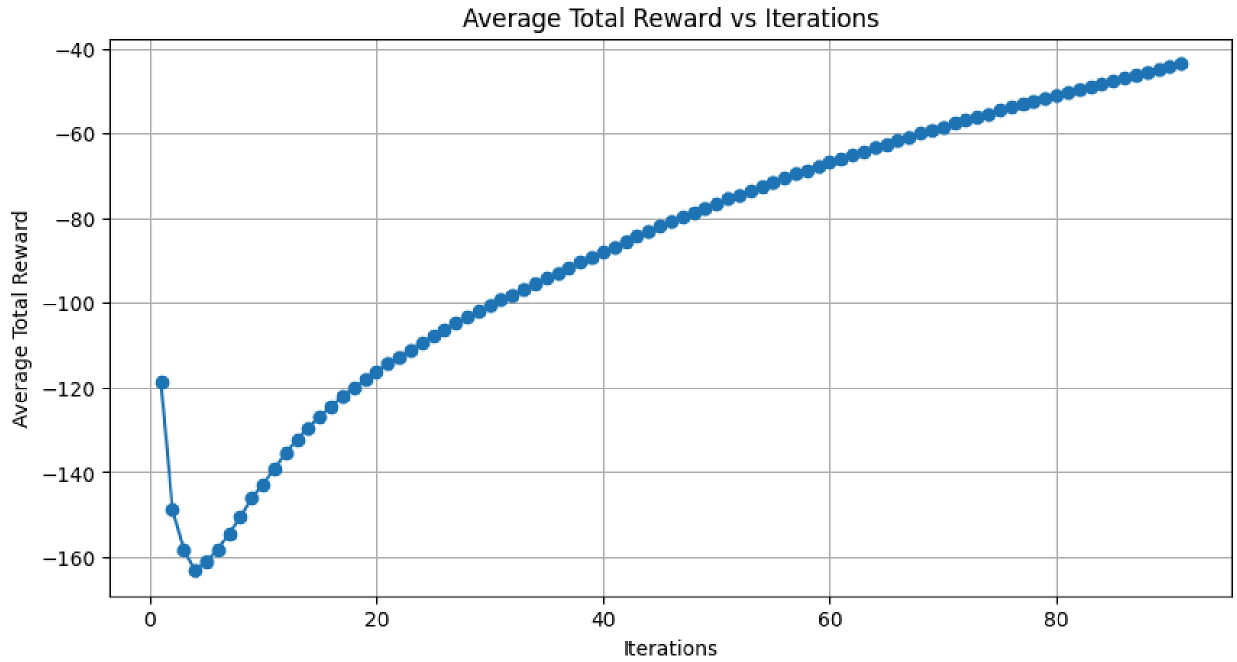


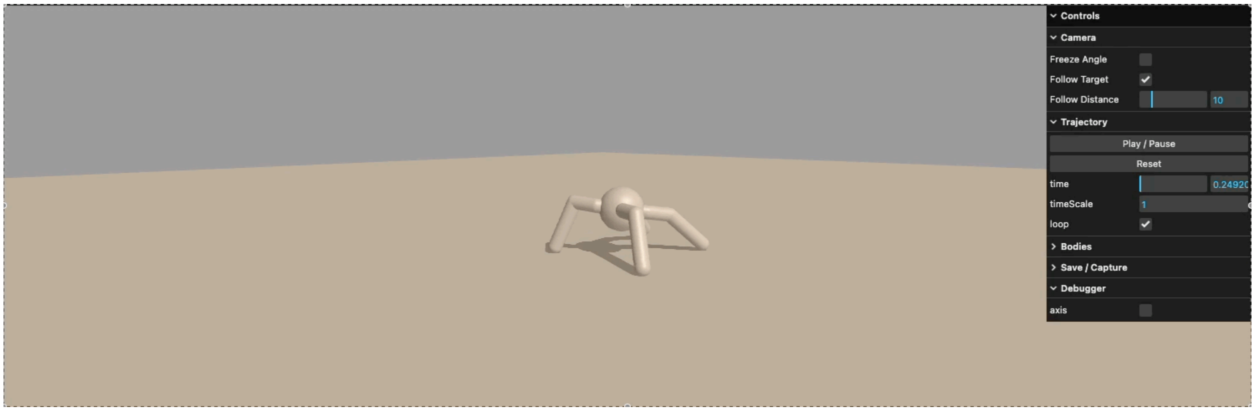


Freeze Angle □
Follow Target ☑
Follow Distance 10
time 0.32290
timeScale 1
loop ☑
axis □

Hyperparameter 3 :

Learning Rate : 3e-5
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.82
Gae Lambda : 0.95
Update Epochs : 10
Clip Coef : 0.35
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.0
Vf Coef : 0.5
Max Grad Norm : 0.5
Target Kl : None
Num Mini Batches : 32
Total Timesteps : 3000000
Minibatch Size : 100
Batch Size : 200
Num Iterations : 100



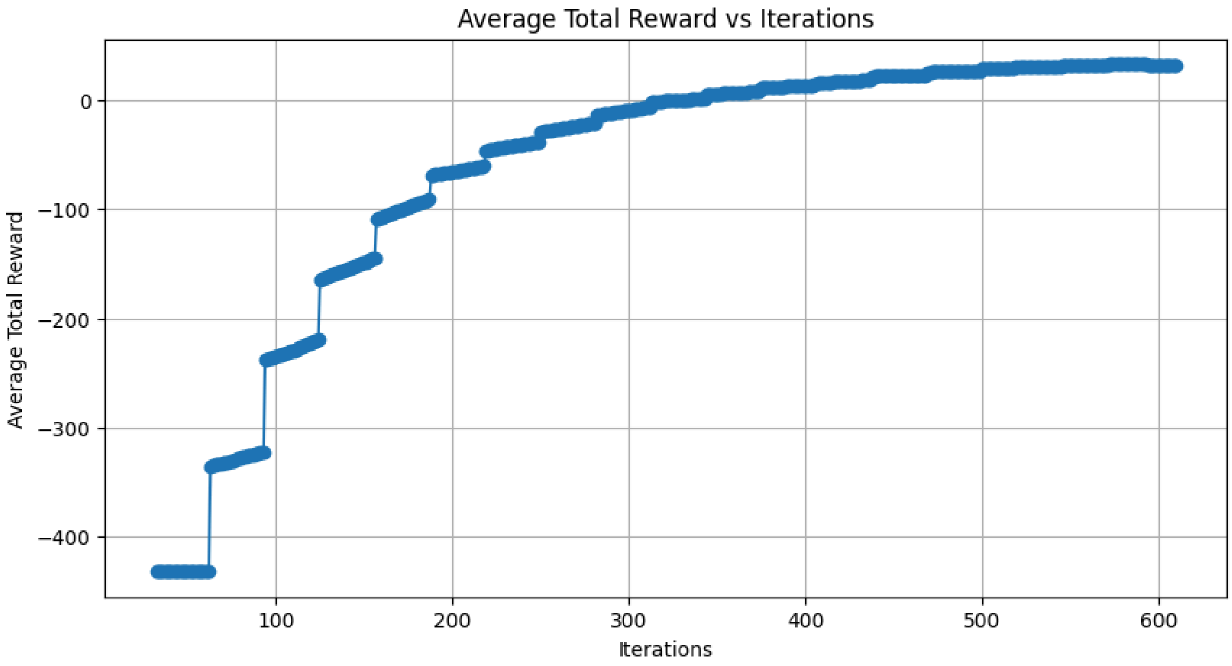


Freeze Angle □
Follow Target ☑
Follow Distance 10
time 0.24920
timeScale 1
loop ☑
axis □

Half Cheetah

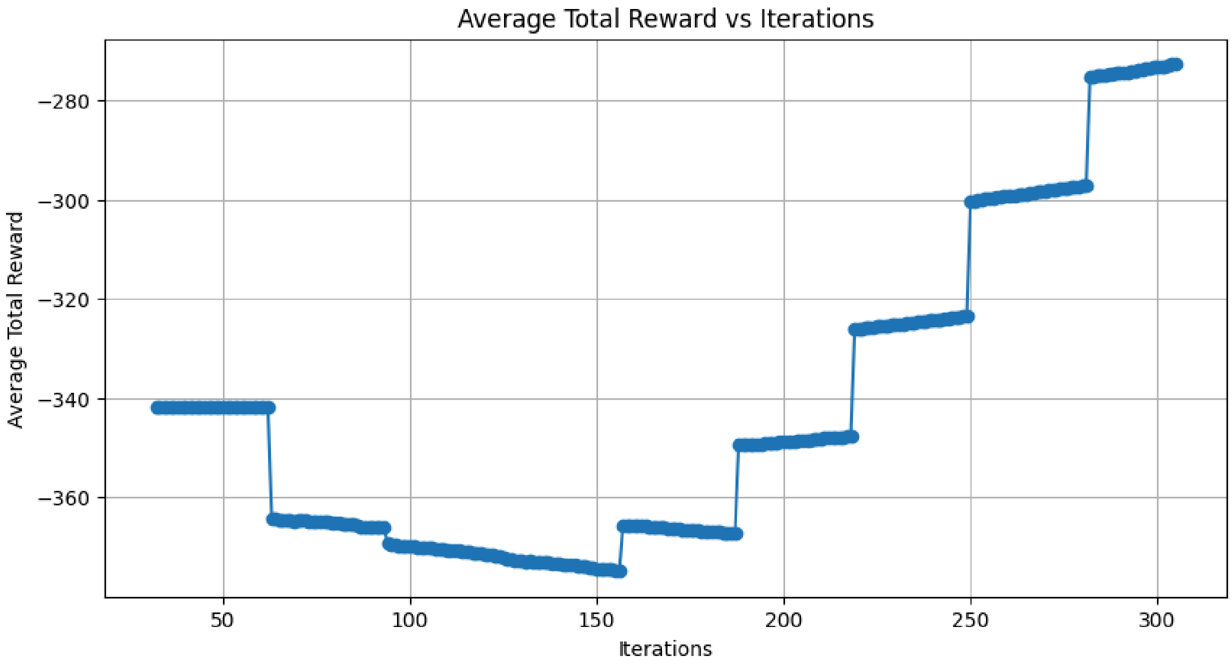
Hyperparameter 1 :

Learning Rate : 5e-5
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.70
Gae Lambda : 0.95
Update Epochs : 10
Clip Coef : 0.35
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.01
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : 0.01
Num Mini Batches : 32
Total Timesteps : 2000000
Minibatch Size : 100
Batch Size : 200
Num Iterations : 600



Hyperparameter 2 :

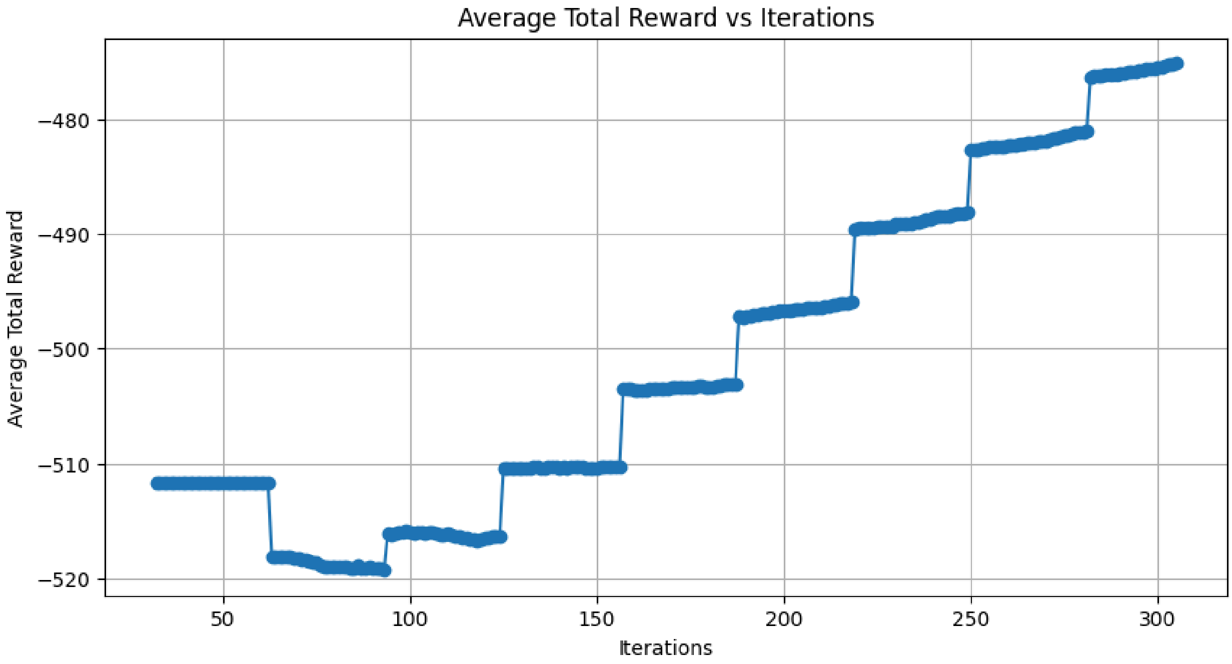
Learning Rate : 1e-5
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.75
Gae Lambda : 0.95
Update Epochs : 10
Clip Coef : 0.40
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.01
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : 0.05
Num Mini Batches : 32
Total Timesteps : 2000000
Minibatch Size : 100
Batch Size : 200
Num Iterations : 300

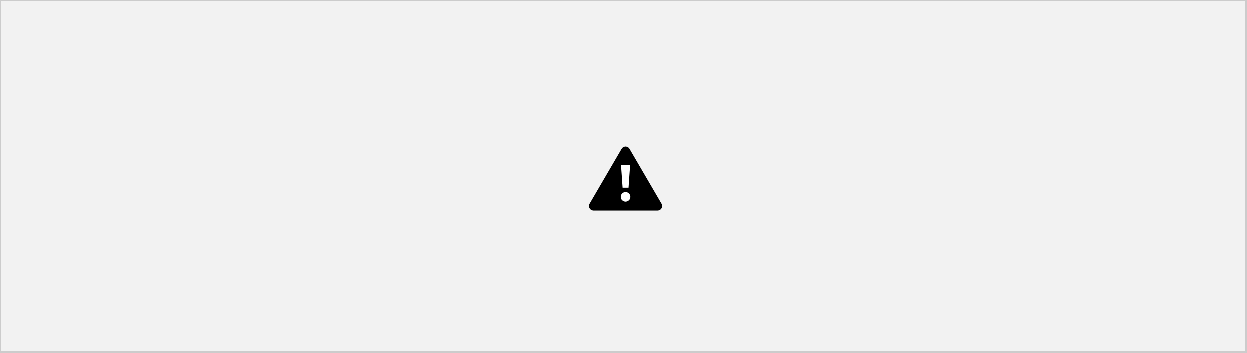


Hyperparameter 3 :

Learning Rate : 3e-7
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.90

Gae Lambda : 0.99
Update Epochs : 10
Clip Coef : 0.35
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.01
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : 0.1
Num Mini Batches : 32
Total Timesteps : 1000000
Minibatch Size : 0
Batch Size : 0
Num Iterations : 400

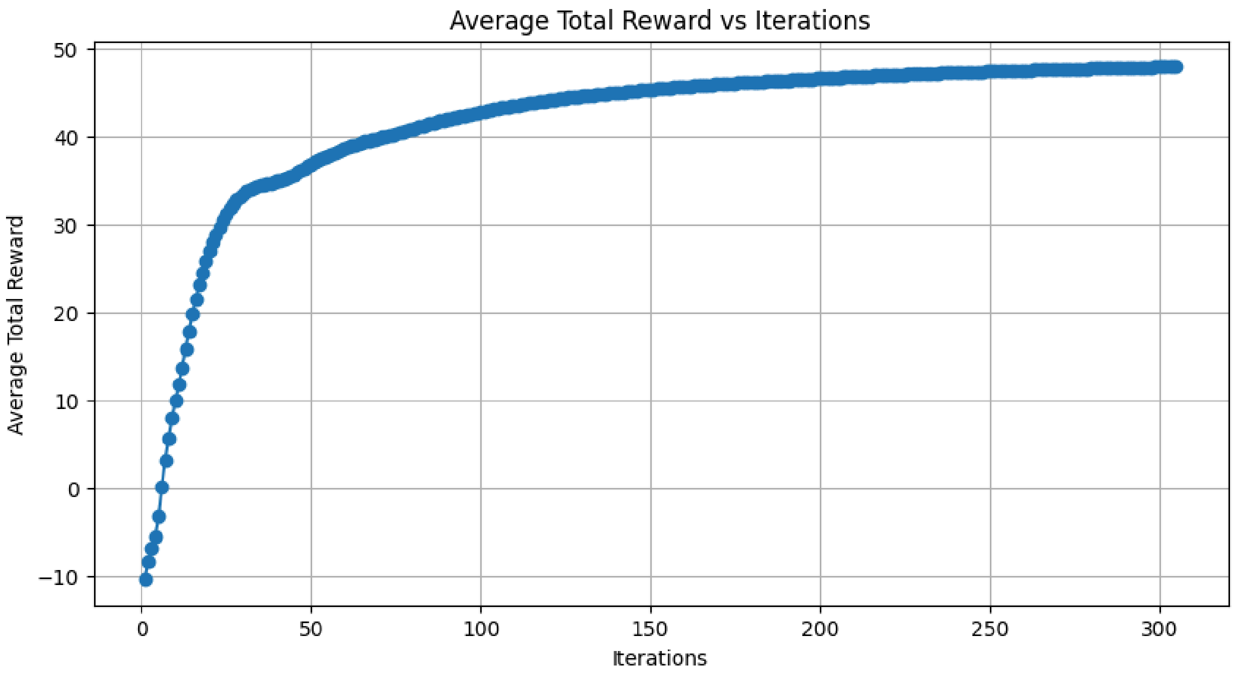




Walker 2D

Hyperparameter 1 :

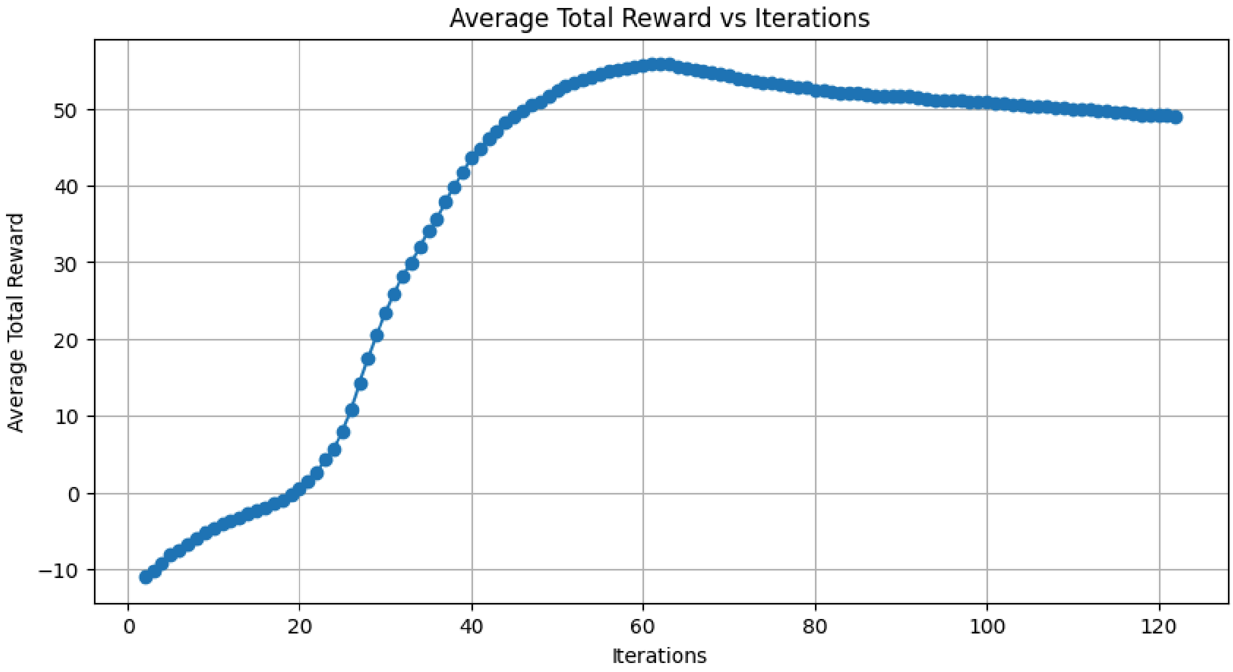
Learning Rate : 3e-4
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.90
Gae Lambda : 0.99
Update Epochs : 10
Clip Coef : 0.40
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.01
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : 0.05
Num Mini Batches : 32
Total Timesteps : 10000000
Minibatch Size : 200
Batch Size : 500
Num Iterations : 400

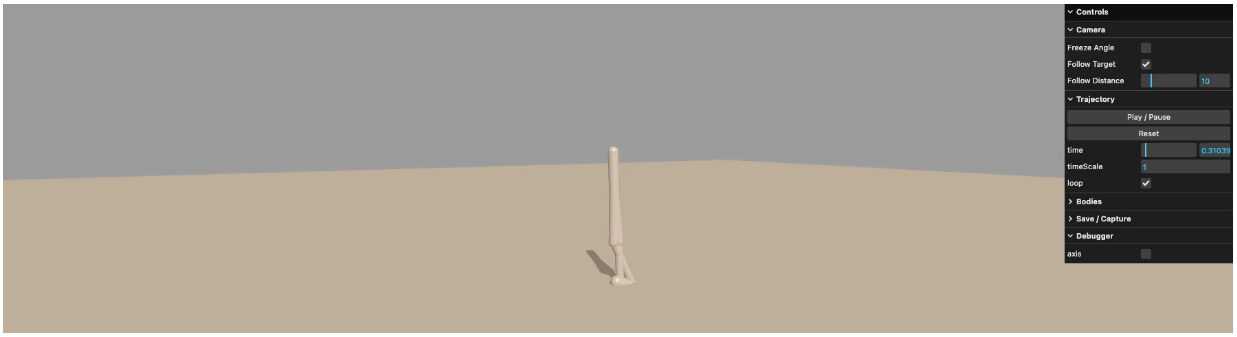


Hyperparameter 2 :

Learning Rate : 1e-4
Num Steps : 8
Num Encs : 1024
Seed : 42
Annela Lr: No
Gamma: 0.85
Gae Lambda : 0.90
Update Epochs : 10
Clip Coef : 0.25
Norm Adv : yes

Clip Vloss:No
Ent Coef : 0.01
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : 0.1
Num Mini Batches : 32
Total Timesteps : 1000000
Minibatch Size : 10
Batch Size : 20
Num Iterations : 150

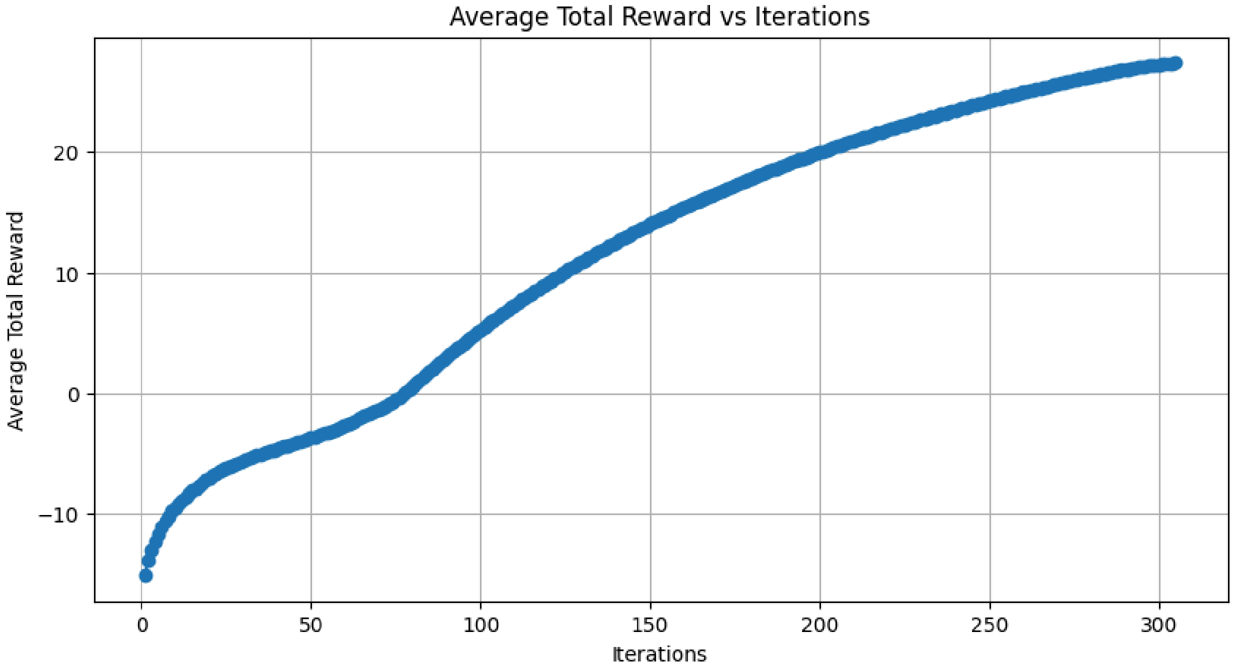


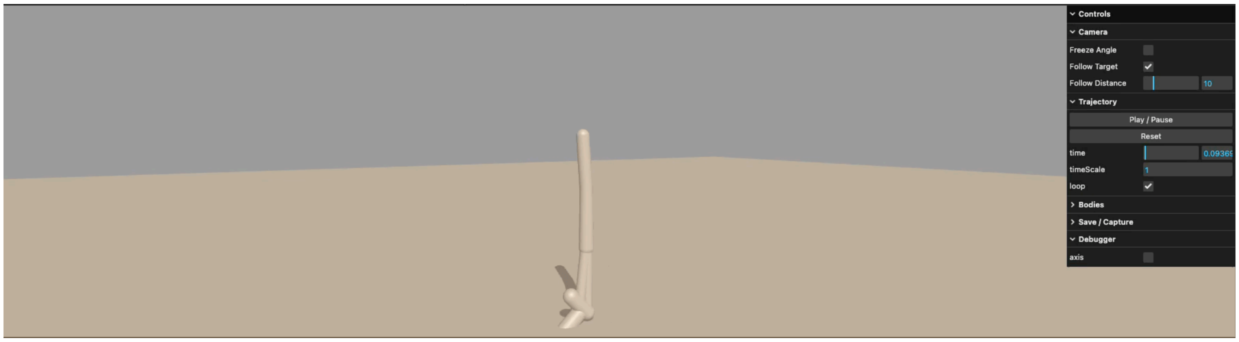


Freeze Angle □
Follow Target ☑
Follow Distance 10
time 0.31039
timeScale 1
loop ☑
axis □

Hyperparameter 3 :

Learning Rate : 3e-5
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.90
Gae Lambda : 0.90
Update Epochs : 10
Clip Coef : 0.40
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.01
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : 0.05
Num Mini Batches : 32
Total Timesteps : 1000000
Minibatch Size : 200
Batch Size : 500
Num Iterations : 400



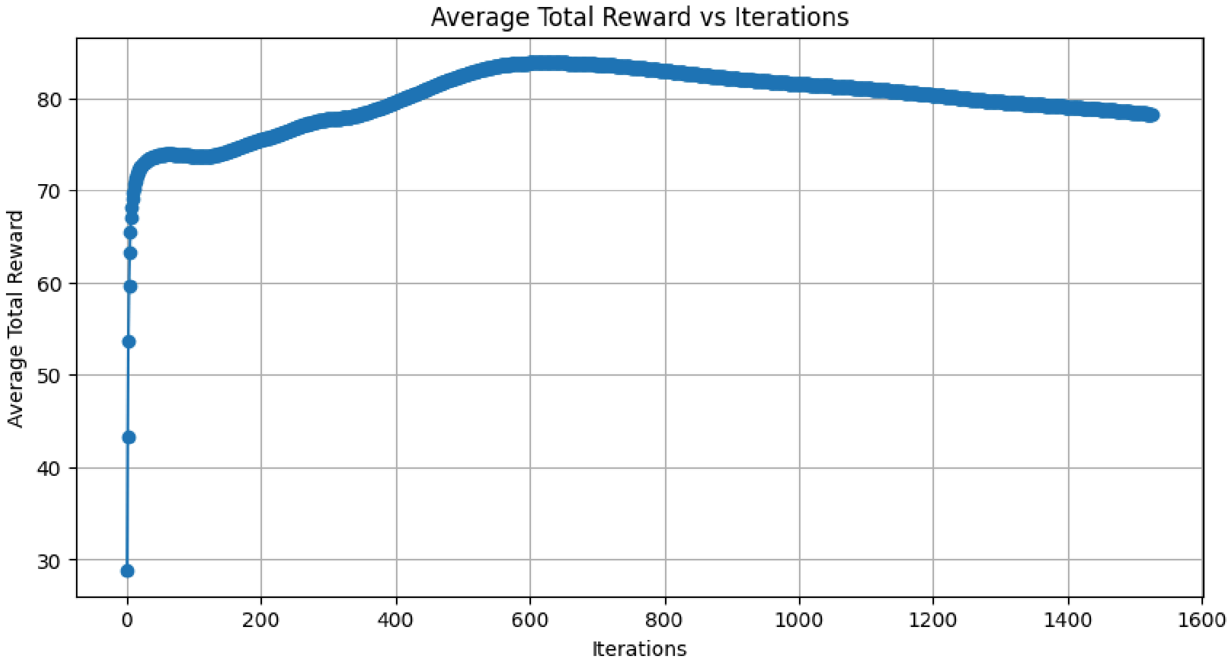


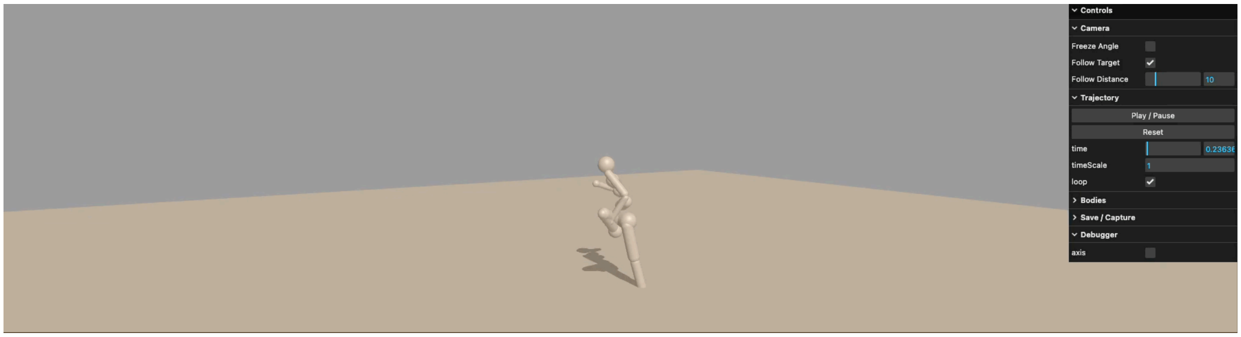
Freeze Angle ☐
Follow Target ☑
Follow Distance 10
time 0.0936
timeScale 1
loop ☑
axis ☐

Humanoid

Hyperparameter 1 :

Learning Rate : 3e-7
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.99
Gae Lambda : 0.99
Update Epochs : 10
Clip Coef : 0.32
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.05
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : None
Num Mini Batches : 32
Total Timesteps : 50000000
Minibatch Size : 100
Batch Size : 250
Num Iterations : 1000



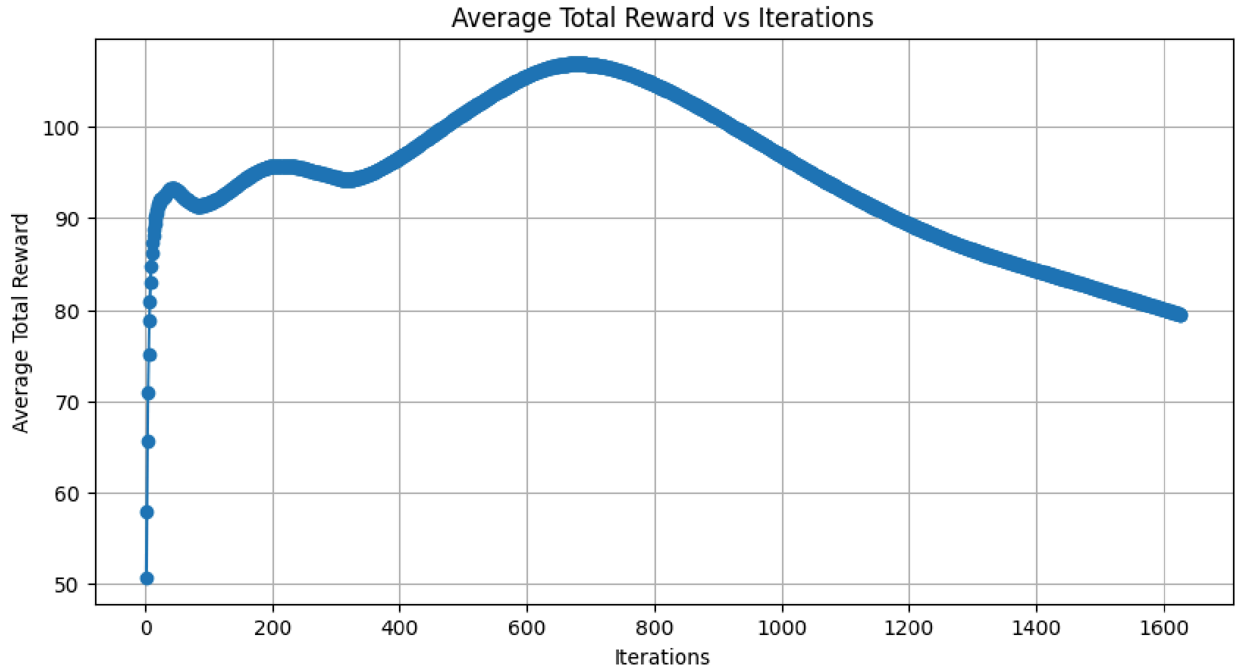


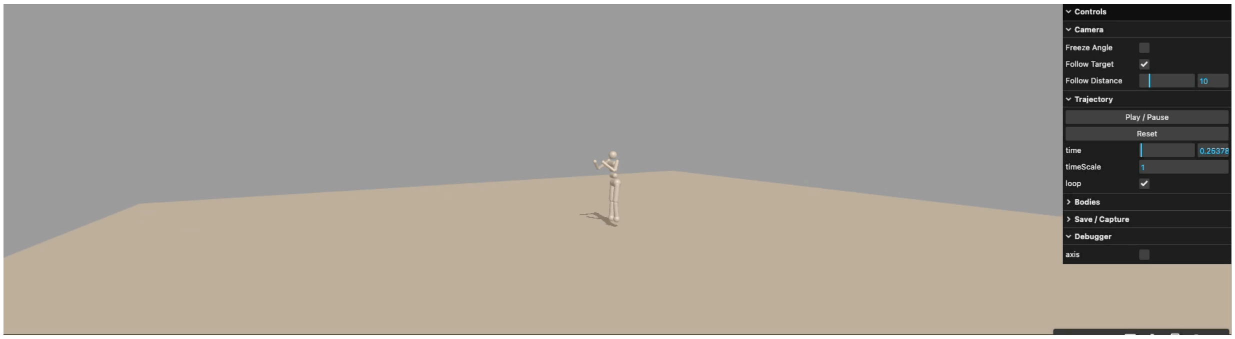
Freeze Angle □
Follow Target ☑
Follow Distance 10
time 0.23636
timeScale 1
loop ☑
axis □

Hyperparameter 2 :

Learning Rate : 3e-5
Num Steps : 16
Num Encs : 2048
Seed : 42
Annela Lr: No
Gamma: 0.85
Gae Lambda : 0.90
Update Epochs : 10
Clip Coef : 0.25
Norm Adv : yes
Clip Vloss:No

Ent Coef : 0.01
Vf Coef : 1.0
Max Grad Norm : 0.5
Target Kl : 0.01
Num Mini Batches : 32
Total Timesteps : 50000000
Minibatch Size : 200
Batch Size : 500
Num Iterations : 1000



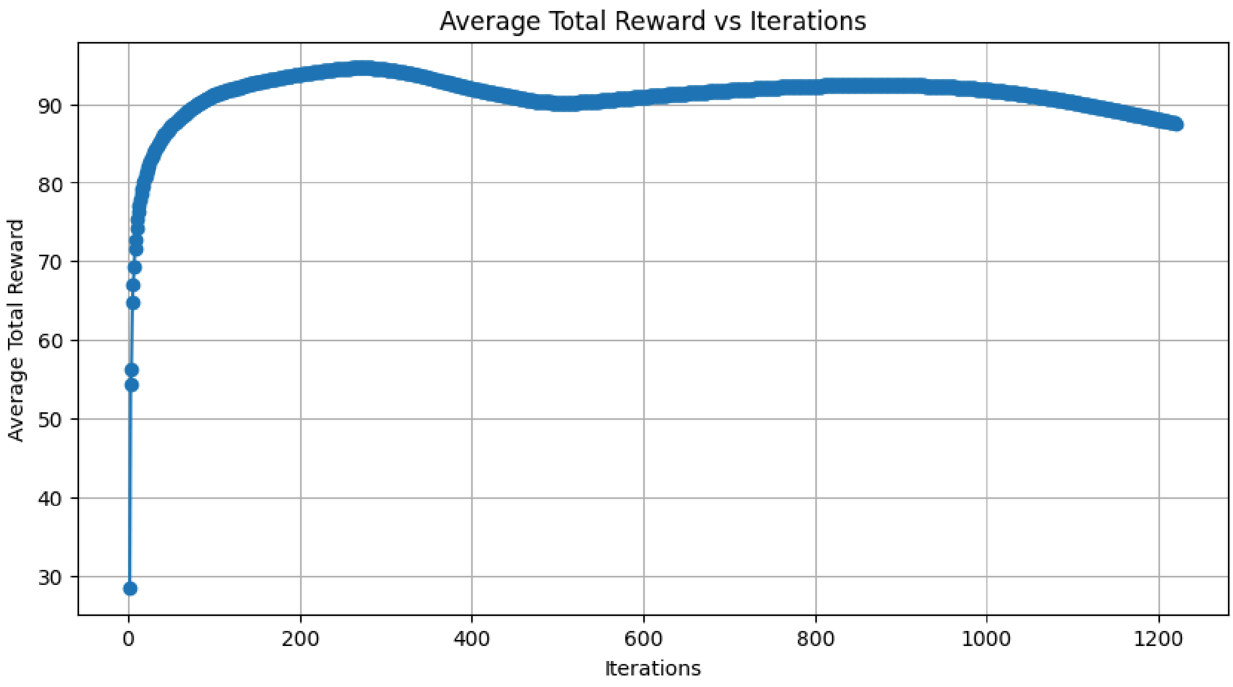


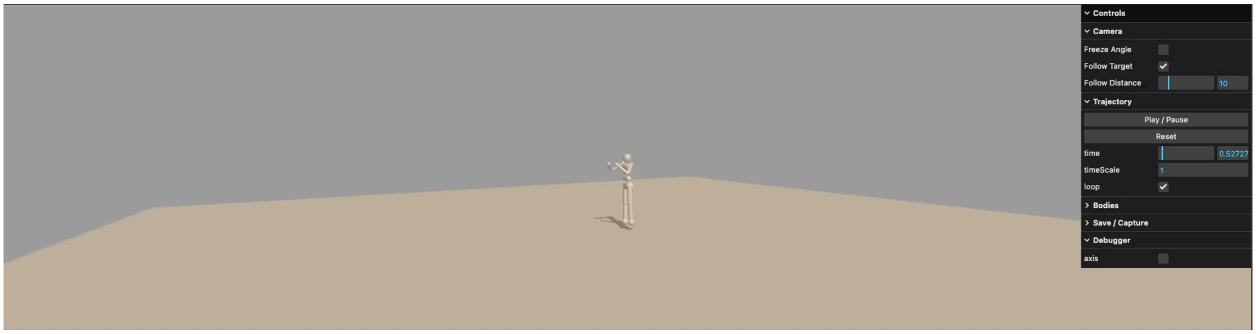
Freeze Angle □
Follow Target ☑
Follow Distance 10
time 0.25378
timeScale 1
loop ☑
axis □

Hyperparameter 3 :

Learning Rate : 1e-4
Num Steps :8
Num Encs : 1024

Seed : 42
Annela Lr: True
Gamma: 0.99
Gae Lambda : 0.95
Update Epochs : 10
Clip Coef : 0.25
Norm Adv : yes
Clip Vloss:No
Ent Coef : 0.0005
Vf Coef : 0.5
Max Grad Norm : 0.5
Target Kl : None
Num Mini Batches : 32
Total Timesteps : 10000000
Minibatch Size : 10
Batch Size : 128
Num Iterations : 100





Freeze Angle ☐
Follow Target ☑
Follow Distance 10
time 0.52727
timeScale 1
loop ☑
axis ☐

Why it is hard to train Humanoid?

1. High-Dimensional Observation and Action Spaces

- **Observation Space**: The humanoid environment has a large observation space that includes joint positions, velocities, angular velocities, and external forces, among other state variables. This dimensionality can easily exceed hundreds of features, making it challenging for reinforcement learning algorithms to find and exploit meaningful patterns.
- **Action Space**: The action space is also high-dimensional, with each joint of the humanoid body (like hips, knees, shoulders) requiring separate control. Each joint must be controlled accurately, meaning the policy must output multiple continuous values for each joint in a way that produces stable, coordinated movement.

2. Explained Variance (explained_var) and Training Challenges

- **Low Explained Variance**: Explained variance is a measure of how well the agent's value function (estimated expected reward) explains the actual rewards received from the environment. In a humanoid environment, the explained variance often remains low because the value function struggles to predict the outcome due to the high-dimensional action space and the chaotic nature of balancing.
- Low explained variance indicates that the value function does not capture the reward trends effectively, often due to excessive noise in the rewards or the value estimates not generalizing well to the observed states. In complex environments like humanoid, high variance in outcomes from similar states or actions makes it difficult for the value function to provide reliable guidance, hindering the learning process.

3. Balance of Exploration and Exploitation

- Due to the difficulty of balancing and moving a humanoid, a large exploration space is required to find successful strategies, which means that the agent must explore a wide range of states and actions before it can exploit known, stable patterns.

- This environment makes it challenging to balance exploration and exploitation because exploration is inherently risky and often leads to failure (e.g., falling), while exploitation of partially learned patterns may not yield further improvements. The agent must continually adjust this balance to avoid falling into local optima where it only partially achieves stable movement.

4. Sensitivity to Hyperparameters

- In humanoid environments, reinforcement learning models can be highly sensitive to hyperparameters, such as learning rate, discount factor, and the frequency of policy updates. Even small changes in these parameters can lead to drastically different behaviors, making tuning a challenging process.
- This sensitivity is because small shifts in how actions are evaluated or how rewards are discounted can significantly impact the balance and stability of the humanoid’s movement.

In summary, the humanoid environment is difficult because it demands high-dimensional, coordinated, and stable control over a complex body with delayed rewards and high sensitivity to initial conditions and hyperparameters. Combined with low explained variance, these factors make the learning process particularly slow and challenging.