Twin Delayed DDPG (TD3)

```
!apt-get update && apt-get install -y xvfb
```

```
Unpacking libxkbfile1:amd64 (1:1.1.0-1build3) ...
Selecting previously unselected package x11-xkb-utils.
Preparing to unpack .../3-x11-xkb-utils_7.7+5build4_amd64.deb ...
Unpacking x11-xkb-utils (7.7+5build4) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../4-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../5-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...
Selecting previously unselected package xfonts-base.
Preparing to unpack .../6-xfonts-base_1%3a1.0.5_all.deb ...
Unpacking xfonts-base (1:1.0.5) ...
Selecting previously unselected package xserver-common.
Preparing to unpack .../7-xserver-common_2%3a21.1.4-2ubuntu1.7~22.04.12_all.deb ...
Unpacking xserver-common (2:21.1.4-2ubuntu1.7~22.04.12) ...
Selecting previously unselected package xvfb.
Preparing to unpack .../8-xvfb_2%3a21.1.4-2ubuntu1.7~22.04.12_amd64.deb ...
Unpacking xvfb (2:21.1.4-2ubuntu1.7~22.04.12) ...
Setting up libfontenc1:amd64 (1:1.1.4-1build3) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) ...
Setting up libxkbfile1:amd64 (1:1.1.0-1build3) ...
Setting up libxfont2:amd64 (1:2.0.5-1build1) ...
Setting up x11-xkb-utils (7.7+5build4) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up xfonts-base (1:1.0.5) ...
Setting up xserver-common (2:21.1.4-2ubuntu1.7~22.04.12) ...
Setting up xvfb (2:21.1.4-2ubuntu1.7~22.04.12) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Processing triggers for libc-bin (2.35-0ubuntu3.4) ...
/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libumf.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libhwloc.so.15 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm_debug.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link
```

```
!pip install swig
```

```
Collecting swig
  Downloading swig-4.2.1.post0-py2.py3-none-manylinux_2_5_x86_64.manylinux1_x86_64.whl.metadata (3.5 kB)
Downloading swig-4.2.1.post0-py2.py3-none-manylinux_2_5_x86_64.manylinux1_x86_64.whl (1.8 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.8/1.8 MB 56.5 MB/s eta 0:00:00
Installing collected packages: swig
Successfully installed swig-4.2.1.post0
```

```
!pip install gym[box2d]==0.23.1
```

```
Collecting gym==0.23.1 (from gym[box2d]==0.23.1)
  Downloading gym-0.23.1.tar.gz (626 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 626.2/626.2 kB 37.4 MB/s eta 0:00:00
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: numpy>=1.18.0 in /usr/local/lib/python3.10/dist-packages (from gym==0.23.1->gym[box2d]==0.23.1) (1.26
Requirement already satisfied: cloudpickle>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from gym==0.23.1->gym[box2d]==0.23.1)
Requirement already satisfied: gym-notices>=0.0.4 in /usr/local/lib/python3.10/dist-packages (from gym==0.23.1->gym[box2d]==0.23.1)
Collecting box2d-py==2.3.5 (from gym[box2d]==0.23.1)
```

```
  Downloading box2d-py-2.3.5.tar.gz (374 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 374.4/374.4 kB 31.0 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting pygame==2.1.0 (from gym[box2d]==0.23.1)
  Downloading pygame-2.1.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (9.5 kB)
Downloading pygame-2.1.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (18.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 18.3/18.3 MB 98.2 MB/s eta 0:00:00
Building wheels for collected packages: gym, box2d-py
  Building wheel for gym (pyproject.toml) ... done
  Created wheel for gym: filename=gym-0.23.1-py3-none-any.whl size=701345 sha256=c4ce2a8ffdd8f442cac20fb90aa38a68772727ff122b73b3cf8
  Stored in directory: /root/.cache/pip/wheels/1a/00/fb/fe5cf2860fb9b7bc860e28f00095a1f42c7b726dd6f42d1acc
  Building wheel for box2d-py (setup.py) ... done
  Created wheel for box2d-py: filename=box2d_py-2.3.5-cp310-cp310-linux_x86_64.whl size=2376095 sha256=195a1af9f6c8c0e80444148490212
  Stored in directory: /root/.cache/pip/wheels/db/8f/6a/eaaadf056fba10a98d986f6dce954e6201ba3126926fc5ad9e
Successfully built gym box2d-py
Installing collected packages: box2d-py, pygame, gym
  Attempting uninstall: pygame
    Found existing installation: pygame 2.6.1
    Uninstalling pygame-2.6.1:
      Successfully uninstalled pygame-2.6.1
  Attempting uninstall: gym
    Found existing installation: gym 0.25.2
    Uninstalling gym-0.25.2:
      Successfully uninstalled gym-0.25.2
Successfully installed box2d-py-2.3.5 gym-0.23.1 pygame-2.1.0
```

```
!pip install pytorch_lightning
```

```
Collecting pytorch_lightning
  Downloading pytorch_lightning-2.4.0-py3-none-any.whl.metadata (21 kB)
Requirement already satisfied: torch>=2.1.0 in /usr/local/lib/python3.10/dist-packages (from pytorch_lightning) (2.5.0+cu121)
Requirement already satisfied: tqdm>=4.57.0 in /usr/local/lib/python3.10/dist-packages (from pytorch_lightning) (4.66.6)
Requirement already satisfied: PyYAML>=5.4 in /usr/local/lib/python3.10/dist-packages (from pytorch_lightning) (6.0.2)
Requirement already satisfied: fsspec>=2022.5.0 in /usr/local/lib/python3.10/dist-packages (from fsspec[http]>=2022.5.0->pytorch_lig
Collecting torchmetrics>=0.7.0 (from pytorch_lightning)
  Downloading torchmetrics-1.5.2-py3-none-any.whl.metadata (20 kB)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from pytorch_lightning) (24.1)
Requirement already satisfied: typing-extensions>=4.4.0 in /usr/local/lib/python3.10/dist-packages (from pytorch_lightning) (4.12.2
Collecting lightning-utilities>=0.10.0 (from pytorch_lightning)
  Downloading lightning_utilities-0.11.8-py3-none-any.whl.metadata (5.2 kB)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in /usr/local/lib/python3.10/dist-packages (from fsspec[http]>=2022.5.0->p
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from lightning-utilities>=0.10.0->pytorch_ligh
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch>=2.1.0->pytorch_lightning) (3.16.1)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch>=2.1.0->pytorch_lightning) (3.4.2)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=2.1.0->pytorch_lightning) (3.1.4)
Requirement already satisfied: sympy==1.13.1 in /usr/local/lib/python3.10/dist-packages (from torch>=2.1.0->pytorch_lightning) (1.13
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from sympy==1.13.1->torch>=2.1.0->pyto
Requirement already satisfied: numpy>1.20.0 in /usr/local/lib/python3.10/dist-packages (from torchmetrics>=0.7.0->pytorch_lightning
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1-
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[htt
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fssp
Requirement already satisfied: yarl<2.0,>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec
Requirement already satisfied: async-timeout<5.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1-
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch>=2.1.0->pytorch_lightn
Requirement already satisfied: idna>=2.0 in /usr/local/lib/python3.10/dist-packages (from yarl<2.0,>=1.12.0->aiohttp!=4.0.0a0,!=4.0
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.10/dist-packages (from yarl<2.0,>=1.12.0->aiohttp!=4.0.0a0
Downloading pytorch_lightning-2.4.0-py3-none-any.whl (815 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 815.2/815.2 kB 43.0 MB/s eta 0:00:00
Downloading lightning_utilities-0.11.8-py3-none-any.whl (26 kB)
Downloading torchmetrics-1.5.2-py3-none-any.whl (891 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 891.4/891.4 kB 59.2 MB/s eta 0:00:00
Installing collected packages: lightning-utilities, torchmetrics, pytorch_lightning
Successfully installed lightning-utilities-0.11.8 pytorch_lightning-2.4.0 torchmetrics-1.5.2
```

```
!pip install pyvirtualdisplay
```

```
Collecting pyvirtualdisplay
  Downloading PyVirtualDisplay-3.0-py3-none-any.whl.metadata (943 bytes)
Downloading PyVirtualDisplay-3.0-py3-none-any.whl (15 kB)
Installing collected packages: pyvirtualdisplay
Successfully installed pyvirtualdisplay-3.0
```

```
!pip install brax==0.10.5
```

```
Requirement already satisfied: PyYAML>=5.4.1 in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.10.5) (6.0.2)
Requirement already satisfied: cloudpickle>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from gym->brax==0.10.5) (3.1.0)
Requirement already satisfied: gym-notices>=0.0.4 in /usr/local/lib/python3.10/dist-packages (from gym->brax==0.10.5) (0.0.8)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from ml-collections->brax==0.10.5) (1.16.0)
Collecting contextlib2 (from ml-collections->brax==0.10.5)
  Downloading contextlib2-21.6.0-py2.py3-none-any.whl.metadata (4.1 kB)
Collecting glfw (from mujoco->brax==0.10.5)
  Downloading glfw-2.7.0-py2.py27.py3.py30.py31.py32.py33.py34.py35.py36.py37.py38-none-manylinux2014_x86_64.whl.metadata (5.4 kB
Requirement already satisfied: pyopengl in /usr/local/lib/python3.10/dist-packages (from mujoco->brax==0.10.5) (3.1.7)
Requirement already satisfied: chex>=0.1.86 in /usr/local/lib/python3.10/dist-packages (from optax->brax==0.10.5) (0.1.87)
Requirement already satisfied: nest_asyncio in /usr/local/lib/python3.10/dist-packages (from orbax-checkpoint->brax==0.10.5) (1.6
Requirement already satisfied: protobuf in /usr/local/lib/python3.10/dist-packages (from orbax-checkpoint->brax==0.10.5) (3.20.3)
Requirement already satisfied: humanize in /usr/local/lib/python3.10/dist-packages (from orbax-checkpoint->brax==0.10.5) (4.11.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorboardX->brax==0.10.5) (24.1)
Requirement already satisfied: toolz>=0.9.0 in /usr/local/lib/python3.10/dist-packages (from chex>=0.1.86->optax->brax==0.10.5) (
Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.10/dist-packages (from rich>=11.1->flax->brax==0.1
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from rich>=11.1->flax->brax==0
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from etils[epath,epy]->orbax-checkpoint->brax==
Requirement already satisfied: importlib_resources in /usr/local/lib/python3.10/dist-packages (from etils[epath,epy]->orbax-check
Requirement already satisfied: zipp in /usr/local/lib/python3.10/dist-packages (from etils[epath,epy]->orbax-checkpoint->brax==0.
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-packages (from markdown-it-py>=2.2.0->rich>=11.1->fla
Downloading brax-0.10.5-py3-none-any.whl (998 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 998.9/998.9 kB 59.9 MB/s eta 0:00:00
Downloading dm_env-1.6-py3-none-any.whl (26 kB)
Downloading Flask_Cors-5.0.0-py2.py3-none-any.whl (14 kB)
Downloading jaxopt-0.8.3-py3-none-any.whl (172 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 172.3/172.3 kB 18.5 MB/s eta 0:00:00
Downloading mujoco-3.2.5-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.3/6.3 MB 116.3 MB/s eta 0:00:00
Downloading mujoco_mjx-3.2.5-py3-none-any.whl (6.7 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.7/6.7 MB 115.2 MB/s eta 0:00:00
Downloading pytinyrenderer-0.0.14-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.9 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.9/1.9 MB 79.0 MB/s eta 0:00:00
Downloading tensorboardX-2.6.2.2-py2.py3-none-any.whl (101 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 101.7/101.7 kB 11.3 MB/s eta 0:00:00
Downloading trimesh-4.5.2-py3-none-any.whl (704 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 704.4/704.4 kB 51.1 MB/s eta 0:00:00
Downloading contextlib2-21.6.0-py2.py3-none-any.whl (13 kB)
Downloading glfw-2.7.0-py2.py27.py3.py30.py31.py32.py33.py34.py35.py36.py37.py38-none-manylinux2014_x86_64.whl (211 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.8/211.8 kB 21.1 MB/s eta 0:00:00
Building wheels for collected packages: ml-collections
  Building wheel for ml-collections (setup.py) ... done
  Created wheel for ml-collections: filename=ml_collections-0.1.1-py3-none-any.whl size=94508 sha256=0b51315a1c1bc3e1259e59cd1d71
  Stored in directory: /root/.cache/pip/wheels/7b/89/c9/a9b87790789e94aadcfc393c283e3ecd5ab916aed0a31be8fe
Successfully built ml-collections
Installing collected packages: pytinyrenderer, glfw, trimesh, tensorboardX, dm-env, contextlib2, ml-collections, mujoco, flask-co
Successfully installed brax-0.10.5 contextlib2-21.6.0 dm-env-1.6 flask-cors-5.0.0 glfw-2.7.0 jaxopt-0.8.3 ml-collections-0.1.1 mu
```

Setup virtual display

```
from pyvirtualdisplay import Display
Display(visible=False, size=(1400, 900)).start()
```

```
<pyvirtualdisplay.display.Display at 0x7f367cca11b0>
```

Import the necessary code libraries

```
import copy
import gym
import torch
import random
import functools
import itertools

import numpy as np
import torch.nn.functional as F

from collections import deque, namedtuple
from IPython.display import HTML
from base64 import b64encode

from torch import nn
from torch.utils.data import DataLoader
from torch.utils.data.dataset import IterableDataset
from torch.optim import AdamW

from pytorch_lightning import LightningModule, Trainer

import brax
from brax import envs
from brax.envs.wrappers import gym as gym_wrapper
```

```
from brax.envs.wrappers import torch as torch_wrapper

from brax.io import html

device = 'cuda' if torch.cuda.is_available() else 'cpu'
num_gpus = torch.cuda.device_count()
```

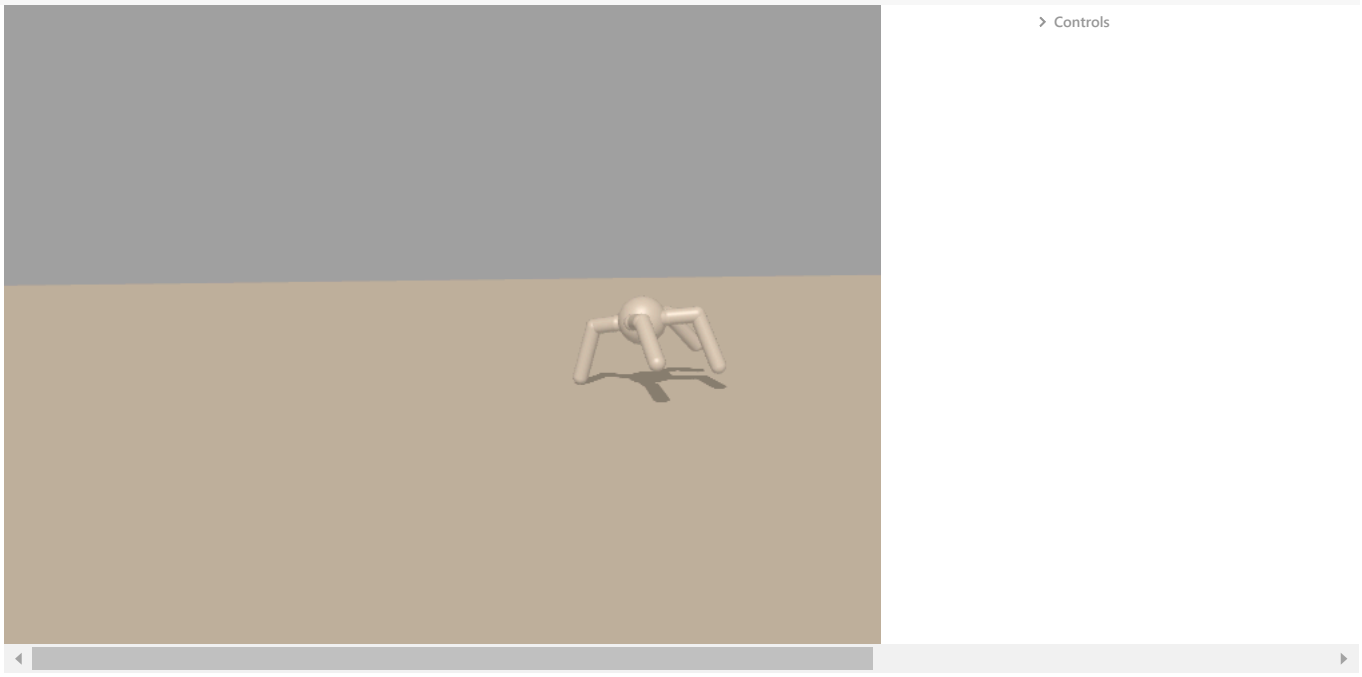
```
def display_video(episode=0):
    video_file = open(f'/content/videos/rl-video-episode-{episode}.mp4', "r+b").read()
    video_url = f"data:video/mp4;base64,{b64encode(video_file).decode()}"
    return HTML(f"<video width=600 controls><source src='{video_url}'></video>")
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_c
  and should_run_async(code)
```

```
def create_environment(env_name, num_envs=256, episode_length=1000):
    env = envs.create(env_name, batch_size=num_envs, episode_length=episode_length, backend='spring')
    env = gym_wrapper.VectorGymWrapper(env)
    env = torch_wrapper.TorchWrapper(env, device=device)
    return env
```

```
@torch.no_grad()
def test_env(env_name, policy=None):
    env = envs.create(env_name, episode_length=1000, backend='spring')
    env = gym_wrapper.GymWrapper(env)
    env = torch_wrapper.TorchWrapper(env, device=device)
    ps_array = []
    state = env.reset()
    for i in range(1000):
        if policy:
            action = algo.policy.net(state.unsqueeze(0)).squeeze()
        else:
            action = torch.from_numpy(env.action_space.sample()).to(device)
        state, _, _, _ = env.step(action)
        ps_array.extend([env.unwrapped._state.pipeline_state]*5)
    return HTML(html.render(env.unwrapped._env.sys, ps_array))
```

```
test_env('ant')
```



Create the gradient policy

```
class GradientPolicy(nn.Module):

    def __init__(self, hidden_size, obs_size, out_dims, min, max):
        super().__init__()
        self.min = torch.from_numpy(min).to(device)
        self.max = torch.from_numpy(max).to(device)
        self.net = nn.Sequential(
            nn.Linear(obs_size, hidden_size),
            nn.ReLU(),
```

```
            nn.Linear(hidden_size, hidden_size),
            nn.ReLU(),
            nn.Linear(hidden_size, out_dims),
            nn.Tanh()
        )

    def mu(self, x):
        if isinstance(x, np.ndarray):
            x = torch.from_numpy(x).to(device)
        return self.net(x.float()) * self.max

    def forward(self, x, epsilon=0.0, noise_clip=None):
        mu = self.mu(x)
        noise = torch.normal(0, epsilon, mu.size(), device=mu.device)
        if noise_clip is not None:
            noise = torch.clamp(noise, - noise_clip, noise_clip)
        mu = mu + noise
        action = torch.max(torch.min(mu, self.max), self.min)
        return action
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_c
  and should_run_async(code)
```

Create the Deep Q-Network

```
class DQN(nn.Module):

    def __init__(self, hidden_size, obs_size, out_dims):
        super().__init__()
        self.net = nn.Sequential(
            nn.Linear(obs_size + out_dims, hidden_size),
            nn.ReLU(),
            nn.Linear(hidden_size, hidden_size),
            nn.ReLU(),
            nn.Linear(hidden_size, 1),
        )

    def forward(self, state, action):
        if isinstance(state, np.ndarray):
            state = torch.from_numpy(state).to(device)
        if isinstance(action, np.ndarray):
            action = torch.from_numpy(action).to(device)
        in_vector = torch.hstack((state, action))
        return self.net(in_vector.float())
```

```
class ReplayBuffer:

    def __init__(self, capacity):
        self.buffer = deque(maxlen=capacity)

    def __len__(self):
        return len(self.buffer)

    def append(self, experience):
        self.buffer.append(experience)

    def sample(self, batch_size):
        return random.sample(self.buffer, batch_size)
```

```
class RLDataset(IterableDataset):

    def __init__(self, buffer, sample_size=400):
        self.buffer = buffer
        self.sample_size = sample_size

    def __iter__(self):
        for experience in self.buffer.sample(self.sample_size):
            yield experience
```

```
def polyak_average(net, target_net, tau=0.01):
    for qp, tp in zip(net.parameters(), target_net.parameters()):
        tp.data.copy_(tau * qp.data + (1 - tau) * tp.data)
```

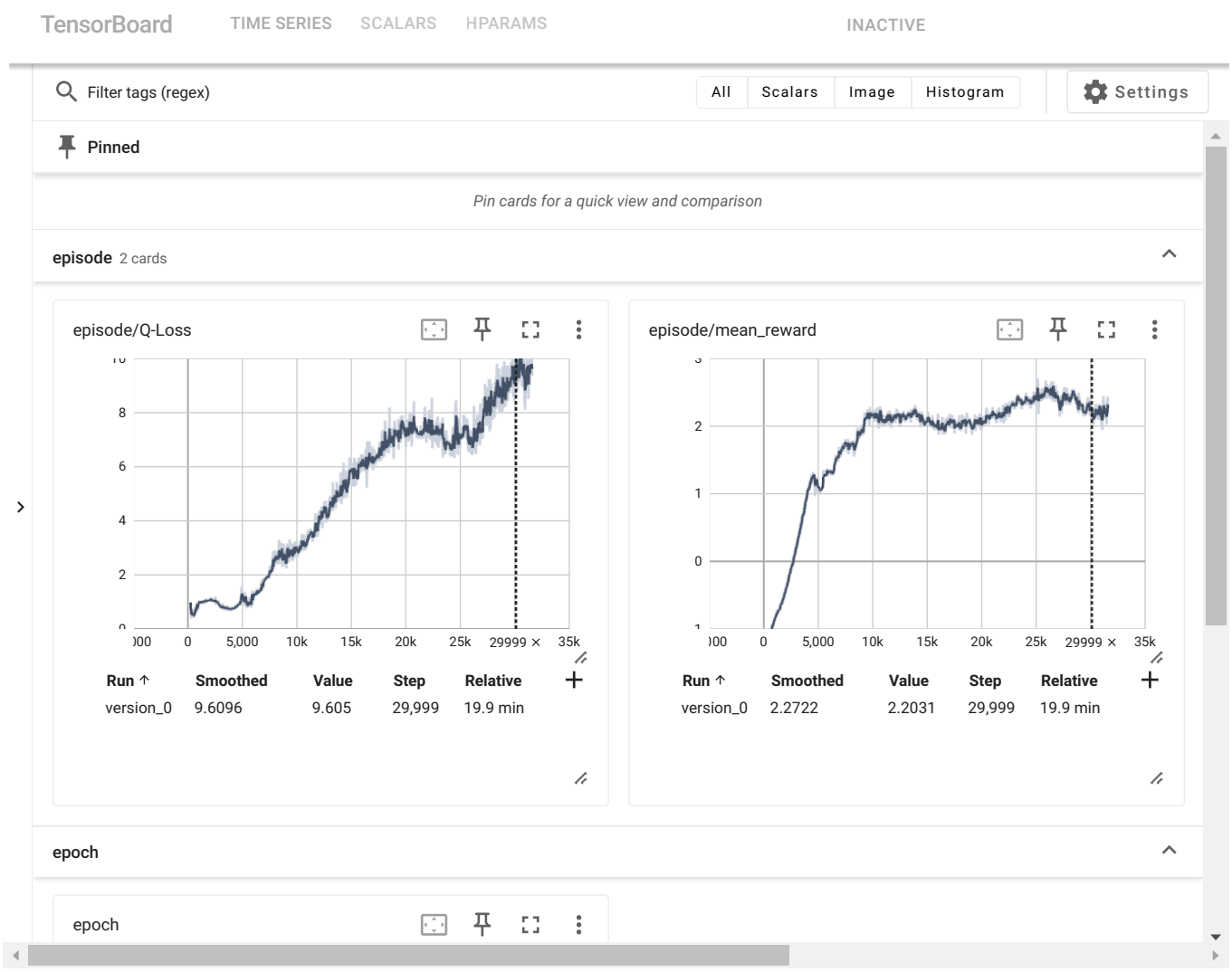
Create the Deep Q-Learning

```
class TD3(LightningModule):

    def __init__(self, env_name, capacity=500, batch_size=8192, actor_lr=1e-3,
                 critic_lr=1e-3, hidden_size=256, gamma=0.99, loss_fn=F.smooth_l1_loss,
                 optim=AdamW, eps_start=1.0, eps_end=0.2, eps_last_episode=500,
                 samples_per_epoch=10, tau=0.005):

        super().__init__()

        self.env = create_environment(env_name, num_envs=batch_size)
        self.obs = self.env.reset()
        self.videos = []

        obs_size = self.env.observation_space.shape[1]
        action_dims = self.env.action_space.shape[1]
        max_action = self.env.action_space.high
        min_action = self.env.action_space.low


        self.q_net1 = DQN(hidden_size, obs_size, action_dims).to(device)
        self.q_net2 = DQN(hidden_size, obs_size, action_dims).to(device)
        self.policy = GradientPolicy(hidden_size, obs_size, action_dims, min_action, max_action).to(device)

        self.target_policy = copy.deepcopy(self.policy)
        self.target_q_net1 = copy.deepcopy(self.q_net1)
        self.target_q_net2 = copy.deepcopy(self.q_net2)

        self.buffer = ReplayBuffer(capacity=capacity)

        self.save_hyperparameters()

        self.automatic_optimization = False

        while len(self.buffer) < self.hparams.samples_per_epoch:
            print(f"{len(self.buffer)} samples in experience buffer. Filling...")
            self.play(epsilon=self.hparams.eps_start)

    @torch.no_grad()
    def play(self, policy=None, epsilon=0.):
        if policy:
            action = policy(self.obs, epsilon=epsilon)
        else:
            action = torch.from_numpy(self.env.action_space.sample()).to(device)
        next_obs, reward, done, info = self.env.step(action)
        exp = (self.obs, action, reward, done, next_obs)
        self.buffer.append(exp)
        self.obs = next_obs
        return reward.mean()

    def forward(self, x):
        output = self.policy.mu(x)
        return output

    def configure_optimizers(self):
        q_net_parameters = itertools.chain(self.q_net1.parameters(), self.q_net2.parameters())
        q_net_optimizer = self.hparams.optim(q_net_parameters, lr=self.hparams.critic_lr)
        policy_optimizer = self.hparams.optim(self.policy.parameters(), lr=self.hparams.actor_lr)
        return [q_net_optimizer, policy_optimizer]

    def train_dataloader(self):
        dataset = RLDataset(self.buffer, self.hparams.samples_per_epoch)
        dataloader = DataLoader(
            dataset=dataset,
            batch_size=1
        )
        return dataloader

    def training_step(self, batch, batch_idx):
        epsilon = max(
            self.hparams.eps_end,
            self.hparams.eps_start - self.current_epoch / self.hparams.eps_last_episode
        )

        mean_reward = self.play(policy=self.policy, epsilon=epsilon)
        self.log('episode/mean_reward', mean_reward)

        polyak_average(self.q_net1, self.target_q_net1, tau=self.hparams.tau)
        polyak_average(self.q_net2, self.target_q_net2, tau=self.hparams.tau)
        polyak_average(self.policy, self.target_policy, tau=self.hparams.tau)

        states, actions, rewards, dones, next_states = map(torch.squeeze, batch)
```

```
        rewards = rewards.unsqueeze(1)
        dones = dones.unsqueeze(1).bool()

        # Optimize critic networks (optimizer_idx 0)
        opt_q, opt_policy = self.optimizers()  # Access optimizers
        opt_q.zero_grad()
        action_values1 = self.q_net1(states, actions)
        action_values2 = self.q_net2(states, actions)
        next_actions = self.target_policy(next_states, epsilon=epsilon, noise_clip=0.05)
        next_action_values = torch.min(
            self.target_q_net1(next_states, next_actions),
            self.target_q_net2(next_states, next_actions),
        )
        next_action_values[dones] = 0.0
        expected_action_values = rewards + self.hparams.gamma * next_action_values
        q_loss1 = self.hparams.loss_fn(action_values1, expected_action_values)
        q_loss2 = self.hparams.loss_fn(action_values2, expected_action_values)
        total_loss = q_loss1 + q_loss2
        self.manual_backward(total_loss)  # Manually backpropagate
        opt_q.step()
        self.log("episode/Q-Loss", total_loss)

        # Optimize policy network (optimizer_idx 1) every 2 steps
        if batch_idx % 2 == 0:
            opt_policy.zero_grad()  # Zero policy gradients
            mu = self.policy.mu(states)
            policy_loss = - self.q_net1(states, mu).mean()
            self.manual_backward(policy_loss)  # Manually backpropagate
            opt_policy.step()  # Update policy network parameters
            self.log

    def on_train_epoch_end(self):
        """This method is called when the training epoch ends."""
        if self.current_epoch % 1000 == 0:
            video = test_env('ant', policy=self.policy)
            self.videos.append(video)
```

```
# Start tensorboard.
!rm -r /content/lightning_logs/
!rm -r /content/videos/
%load_ext tensorboard
%tensorboard --logdir /content/lightning_logs/
```

```
rm: cannot remove '/content/lightning_logs/': No such file or directory
rm: cannot remove '/content/videos/': No such file or directory
```




```
import pytorch_lightning as pl
import warnings
warnings.filterwarnings('ignore')
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_c
  and should_run_async(code)
```

```
algo = TD3('ant')

trainer = pl.Trainer(
    accelerator="gpu" if num_gpus else "cpu", # Use 'gpu' if num_gpus is greater than 0, otherwise use 'cpu'
    devices=1, # Specify the number of GPUs or 'auto' for automatic detection
    max_epochs=3000,
    log_every_n_steps=10
)

trainer.fit(algo)
```

```
0 samples in experience buffer. Filling...
1 samples in experience buffer. Filling...
2 samples in experience buffer. Filling...
3 samples in experience buffer. Filling...
4 samples in experience buffer. Filling...
5 samples in experience buffer. Filling...
6 samples in experience buffer. Filling...
7 samples in experience buffer. Filling...
8 samples in experience buffer. Filling...
INFO:pytorch_lightning.utilities.rank_zero:GPU available: True (cuda), used: True
INFO:pytorch_lightning.utilities.rank_zero:TPU available: False, using: 0 TPU cores
INFO:pytorch_lightning.utilities.rank_zero:HPU available: False, using: 0 HPUs
9 samples in experience buffer. Filling...
INFO:pytorch_lightning.accelerators.cuda:LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]
INFO:pytorch_lightning.callbacks.model_summary:
```

```
algo.videos[2]
```

Controls