

✓ REINFORCE for continuous action spaces

```
!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/libhwloc.so.15 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/libtbbmalloc.so.2 is not a symbolic link

/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/libtbbbind_2_5.so.3 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/libtcm.so.1 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/libtbb.so.12 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/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_openc1.so.0 is not a symbolic link

```

```
!pip install gym==0.23.1
```

```

Collecting gym==0.23.1
  Downloading gym-0.23.1.tar.gz (626 kB)
    626.2/626.2 kB 17.1 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) (1.26.4)
Requirement already satisfied: cloudpickle>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from gym==0.23.1) (3.1.0)
Requirement already satisfied: gym_notices>=0.0.4 in /usr/local/lib/python3.10/dist-packages (from gym==0.23.1) (0.0.8)
Building wheels for collected packages: gym
  Building wheel for gym (pyproject.toml) ... done
  Created wheel for gym: filename=gym-0.23.1-py3-none-any.whl size=701371 sha256=34040a36deddfc1bb5250fb5cef1e14643589d7f41d8c7110d5
  Stored in directory: /root/.cache/pip/wheels/1a/00/fb/fe5cf2860fb9b7bc860e28f00095a1f42c7b726dd6f42d1acc
Successfully built gym
Installing collected packages: gym
  Attempting uninstall: gym
    Found existing installation: gym 0.25.2
    Uninstalling gym-0.25.2:
      Successfully uninstalled gym-0.25.2
  Successfully installed gym-0.23.1

```

```
!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) (2.5.1+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-lightning) (2022.11.0)
Collecting torchmetrics>=0.7.0 (from pytorch-lightning)
  Downloading torchmetrics-1.6.0-py3-none-any.whl.metadata (20 kB)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from pytorch-lightning) (24.2)
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-0.11.9-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->pytorch-lightning) (3.10.10)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from lightning-utilities>=0.10.0->pytorch-lightning) (68.0.0)
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) (1.13.1)
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->pytorch-lightning) (1.3.0)
Requirement already satisfied: numpy>1.20.0 in /usr/local/lib/python3.10/dist-packages (from torchmetrics>=0.7.0->pytorch-lightning) (2.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (2.4.4)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (1.3.1)
Requirement already satisfied: async-timeout<6.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (4.0.3)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (25.1.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (1.4.1)
Requirement already satisfied: multidict>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (6.0.5)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (0.2.0)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (1.17.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2->torch>=2.1.0->pytorch-lightning) (3.0.2)
Requirement already satisfied: idna>=2.0 in /usr/local/lib/python3.10/dist-packages (from yarl<2.0,>=1.17.0->aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]>=2022.5.0->pytorch-lightning) (3.10.3)
Downloading pytorch_lightning-2.4.0-py3-none-any.whl (815 kB)
815.2/815.2 kB 26.8 MB/s eta 0:00:00
Downloading lightning_utilities-0.11.9-py3-none-any.whl (28 kB)
Downloading torchmetrics-1.6.0-py3-none-any.whl (926 kB)
926.4/926.4 kB 38.5 MB/s eta 0:00:00
Installing collected packages: lightning-utilities, torchmetrics, pytorch-lightning
Successfully installed lightning-utilities-0.11.9 pytorch-lightning-2.4.0 torchmetrics-1.6.0
```

```
!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 -U brax==0.0.13 jax==0.3.14 jaxlib==0.3.14+cuda11.cudnn82 -f https://storage.googleapis.com/jax-releases/jax_cuda_releases.html
```

Looking in links: https://storage.googleapis.com/jax-releases/jax_cuda_releases.html
Collecting brax==0.0.13
 Downloading brax-0.0.13-py3-none-any.whl.metadata (5.4 kB)
Collecting jax==0.3.14
 Downloading jax-0.3.14.tar.gz (990 kB)
----- 990.1/990.1 kB 23.0 MB/s eta 0:00:00
Preparing metadata (setup.py) ... done
Collecting jaxlib==0.3.14+cuda11.cudnn82
 Downloading https://storage.googleapis.com/jax-releases/cuda11/jaxlib-0.3.14%2Bcuda11.cudnn82-cp310-none-manylinux2014_x86_64.whl
----- 161.9/161.9 MB 5.6 MB/s eta 0:00:00
Requirement already satisfied: absl-py in /usr/local/lib/python3.10/dist-packages (from brax==0.0.13) (1.4.0)
Collecting dataclasses (from brax==0.0.13)
 Downloading dataclasses-0.6-py3-none-any.whl.metadata (3.0 kB)
Requirement already satisfied: flax in /usr/local/lib/python3.10/dist-packages (from brax==0.0.13) (0.8.5)
Requirement already satisfied: gym in /usr/local/lib/python3.10/dist-packages (from brax==0.0.13) (0.23.1)
Requirement already satisfied: grpcio in /usr/local/lib/python3.10/dist-packages (from brax==0.0.13) (1.68.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from brax==0.0.13) (1.26.4)
Requirement already satisfied: optax in /usr/local/lib/python3.10/dist-packages (from brax==0.0.13) (0.2.4)
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from brax==0.0.13) (11.0.0)
Collecting pytinyrenderer (from brax==0.0.13)
 Downloading pytinyrenderer-0.0.14-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (1.2 kB)
Collecting tensorboardX (from brax==0.0.13)
 Downloading tensorboardX-2.6.2.2-py2.py3-none-any.whl.metadata (5.8 kB)
Collecting trimesh (from brax==0.0.13)
 Downloading trimesh-4.5.3-py3-none-any.whl.metadata (18 kB)
Requirement already satisfied: opt_einsum in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (3.4.0)
Requirement already satisfied: scipy>=1.5 in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (1.13.1)
Requirement already satisfied: typing_extensions in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (4.12.2)
Requirement already satisfied: etils[epath] in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (1.11.0)
Collecting flatbuffers<3.0,>=1.12 (from jaxlib==0.3.14+cuda11.cudnn82)
 Downloading flatbuffers-2.0.7-py2.py3-none-any.whl.metadata (872 bytes)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from etils[epath]->jax==0.3.14) (2024.10.0)
Requirement already satisfied: importlib_resources in /usr/local/lib/python3.10/dist-packages (from etils[epath]->jax==0.3.14) (6.4)
Requirement already satisfied: zipp in /usr/local/lib/python3.10/dist-packages (from etils[epath]->jax==0.3.14) (3.21.0)
INFO: pip is looking at multiple versions of flax to determine which version is compatible with other requirements. This could take
Collecting flax (from brax==0.0.13)
 Downloading flax-0.10.2-py3-none-any.whl.metadata (11 kB)
 Downloading flax-0.10.1-py3-none-any.whl.metadata (11 kB)
 Downloading flax-0.10.0-py3-none-any.whl.metadata (11 kB)
 Downloading flax-0.9.0-py3-none-any.whl.metadata (11 kB)
 Downloading flax-0.8.4-py3-none-any.whl.metadata (10 kB)
 Downloading flax-0.8.3-py3-none-any.whl.metadata (10 kB)
 Downloading flax-0.8.2-py3-none-any.whl.metadata (10 kB)
INFO: pip is still looking at multiple versions of flax to determine which version is compatible with other requirements. This could
 Downloading flax-0.8.1-py3-none-any.whl.metadata (10 kB)
 Downloading flax-0.8.0-py3-none-any.whl.metadata (10 kB)
 Downloading flax-0.7.5-py3-none-any.whl.metadata (10 kB)
 Downloading flax-0.7.4-py3-none-any.whl.metadata (10 kB)
 Downloading flax-0.7.2-py3-none-any.whl.metadata (10.0 kB)
INFO: This is taking longer than usual. You might need to provide the dependency resolver with stricter constraints to reduce runtime
 Downloading flax-0.7.0-py3-none-any.whl.metadata (9.9 kB)
 Downloading flax-0.6.11-py3-none-any.whl.metadata (9.9 kB)
 Downloading flax-0.6.10-py3-none-any.whl.metadata (9.9 kB)
 Downloading flax-0.6.9-py3-none-any.whl.metadata (9.9 kB)
 Downloading flax-0.6.8-py3-none-any.whl.metadata (9.9 kB)
 Downloading flax-0.6.7-py3-none-any.whl.metadata (9.8 kB)
 Downloading flax-0.6.6-py3-none-any.whl.metadata (9.6 kB)
 Downloading flax-0.6.4-py3-none-any.whl.metadata (9.6 kB)
 Downloading flax-0.6.3-py3-none-any.whl.metadata (9.5 kB)
 Downloading flax-0.6.2-py3-none-any.whl.metadata (9.5 kB)
 Downloading flax-0.6.1-py3-none-any.whl.metadata (9.4 kB)
 Downloading flax-0.6.0-py3-none-any.whl.metadata (9.4 kB)
 Downloading flax-0.5.3-py3-none-any.whl.metadata (9.4 kB)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.0.13) (3.8.0)
Requirement already satisfied: msgpack in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.0.13) (1.1.0)
Collecting rich~=11.1 (from flax->brax==0.0.13)
 Downloading rich-11.2.0-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: tensorstore in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.0.13) (0.1.69)
Requirement already satisfied: PyYAML>=5.4.1 in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.0.13) (6.0.2)
Requirement already satisfied: cloudpickle>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from gym->brax==0.0.13) (3.1.0)
Requirement already satisfied: gym_notices>=0.0.4 in /usr/local/lib/python3.10/dist-packages (from gym->brax==0.0.13) (0.0.8)
Requirement already satisfied: chech>=0.1.87 in /usr/local/lib/python3.10/dist-packages (from optax->brax==0.0.13) (0.1.87)
INFO: pip is looking at multiple versions of optax to determine which version is compatible with other requirements. This could take
Collecting optax (from brax==0.0.13)
 Downloading optax-0.2.3-py3-none-any.whl.metadata (8.3 kB)
 Downloading optax-0.2.2-py3-none-any.whl.metadata (8.1 kB)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorboardX->brax==0.0.13) (24.2)
Requirement already satisfied: protobuf>=3.20 in /usr/local/lib/python3.10/dist-packages (from tensorboardX->brax==0.0.13) (4.25.5)
INFO: pip is looking at multiple versions of chech to determine which version is compatible with other requirements. This could take
Collecting chech>=0.1.86 (from optax->brax==0.0.13)
 Downloading chech-0.1.88-py3-none-any.whl.metadata (17 kB)
 Downloading chech-0.1.86-py3-none-any.whl.metadata (17 kB)
Collecting optax (from brax==0.0.13)
 Downloading optax-0.2.1-py3-none-any.whl.metadata (8.0 kB)
Collecting chech>=0.1.7 (from optax->brax==0.0.13)
 Downloading chech-0.1.85-py3-none-any.whl.metadata (17 kB)
 Downloading chech-0.1.84-py3-none-any.whl.metadata (17 kB)
INFO: pip is still looking at multiple versions of chech to determine which version is compatible with other requirements. This could
 Downloading chech-0.1.83-py3-none-any.whl.metadata (17 kB)
 Downloading chech-0.1.82-py3-none-any.whl.metadata (17 kB)

```

Downloading chex-0.1.81-py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: dm-tree>=0.1.5 in /usr/local/lib/python3.10/dist-packages (from chex>=0.1.7->optax->brax==0.0.13) (0
Downloading chex-0.1.7-py3-none-any.whl.metadata (17 kB)
Collecting optax (from brax==0.0.13)
  Downloading optax-0.2.0-py3-none-any.whl.metadata (8.0 kB)
INFO: This is taking longer than usual. You might need to provide the dependency resolver with stricter constraints to reduce runtim
INFO: pip is still looking at multiple versions of optax to determine which version is compatible with other requirements. This coul
  Downloading optax-0.1.9-py3-none-any.whl.metadata (6.7 kB)
  Downloading optax-0.1.8-py3-none-any.whl.metadata (14 kB)
  Downloading optax-0.1.7-py3-none-any.whl.metadata (13 kB)
Collecting chex>=0.1.5 (from optax->brax==0.0.13)
  Downloading chex-0.1.6-py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: toolz>=0.9.0 in /usr/local/lib/python3.10/dist-packages (from chex>=0.1.5->optax->brax==0.0.13) (0.1
Collecting colorama<0.5.0,>=0.4.0 (from rich~=11.1->flax->brax==0.0.13)
  Downloading colorama-0.4.6-py2.py3-none-any.whl.metadata (17 kB)
Collecting commonmark<0.10.0,>=0.9.0 (from rich~=11.1->flax->brax==0.0.13)
  Downloading commonmark-0.9.1-py2.py3-none-any.whl.metadata (5.7 kB)
Requirement already satisfied: pygments<3.0.0,>=2.6.0 in /usr/local/lib/python3.10/dist-packages (from rich~=11.1->flax->brax==0.0.1
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax->brax==0.0.13) (1
Requirement already satisfied: cyclo>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax->brax==0.0.13) (0.12.1
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax->brax==0.0.13) (4
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax->brax==0.0.13) (1
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax->brax==0.0.13) (3
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax->brax==0.0.13) (6
Requirement already satisfied: ml-dtypes>=0.3.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow->flax->brax==0.0.13) (6
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib->flax->br
Downloading brax-0.0.13-py3-none-any.whl (300 kB)
 300.1/300.1 kB 16.8 MB/s eta 0:00:00
Downloading flatbuffers-2.0.7-py2.py3-none-any.whl (26 kB)
Downloading dataclasses-0.6-py3-none-any.whl (14 kB)
Downloading flax-0.5.3-py3-none-any.whl (202 kB)
 202.0/202.0 kB 13.2 MB/s eta 0:00:00
Downloading optax-0.1.7-py3-none-any.whl (154 kB)
 154.1/154.1 kB 11.4 MB/s eta 0:00:00
Downloading chex-0.1.6-py3-none-any.whl (87 kB)
 87.9/87.9 kB 5.7 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 34.3 MB/s eta 0:00:00
Downloading tensorboardX-2.6.2.2-py2.py3-none-any.whl (101 kB)
 101.7/101.7 kB 7.0 MB/s eta 0:00:00
Downloading trimesh-4.5.3-py3-none-any.whl (704 kB)
 704.8/704.8 kB 32.1 MB/s eta 0:00:00
Downloading rich-11.2.0-py3-none-any.whl (217 kB)
 217.3/217.3 kB 13.5 MB/s eta 0:00:00
Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
Downloading commonmark-0.9.1-py2.py3-none-any.whl (51 kB)
 51.1/51.1 kB 3.6 MB/s eta 0:00:00
Building wheels for collected packages: jax
  Building wheel for jax (setup.py) ... done
  Created wheel for jax: filename=jax-0.3.14-py3-none-any.whl size=1147563 sha256=577f2cec4e48f1b576572c632e1117a76684ad0780a466c5ff
  Stored in directory: /root/.cache/pip/wheels/cc/b2/fd/3e8a8312ad916a677b3beb844bfba172999a08244fb1395149
Successfully built jax
Installing collected packages: pytinyrenderer, flatbuffers, dataclasses, commonmark, trimesh, tensorboardX, colorama, rich, jaxlib,
  Attempting uninstall: flatbuffers
    Found existing installation: flatbuffers 24.3.25
    Uninstalling flatbuffers-24.3.25:
      Successfully uninstalled flatbuffers-24.3.25
  Attempting uninstall: rich
    Found existing installation: rich 13.9.4
    Uninstalling rich-13.9.4:
      Successfully uninstalled rich-13.9.4
  Attempting uninstall: jaxlib
    Found existing installation: jaxlib 0.4.33
    Uninstalling jaxlib-0.4.33:
      Successfully uninstalled jaxlib-0.4.33
  Attempting uninstall: jax
    Found existing installation: jax 0.4.33
    Uninstalling jax-0.4.33:
      Successfully uninstalled jax-0.4.33
  Attempting uninstall: chex
    Found existing installation: chex 0.1.87
    Uninstalling chex-0.1.87:
      Successfully uninstalled chex-0.1.87
  Attempting uninstall: optax
    Found existing installation: optax 0.2.4
    Uninstalling optax-0.2.4:
      Successfully uninstalled optax-0.2.4
  Attempting uninstall: flax
    Found existing installation: flax 0.8.5
    Uninstalling flax-0.8.5:
      Successfully uninstalled flax-0.8.5
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the sou
ibis-framework 9.2.0 requires rich<14,>=12.4.4, but you have rich 11.2.0 which is incompatible.
orbax-checkpoint 0.6.4 requires jax>=0.4.26, but you have jax 0.3.14 which is incompatible.
pymc 5.18.2 requires rich>=13.7.1, but you have rich 11.2.0 which is incompatible.
tensorflow 2.17.1 requires flatbuffers>=24.3.25, but you have flatbuffers 2.0.7 which is incompatible.
Successfully installed brax-0.0.13 chex-0.1.6 colorama-0.4.6 commonmark-0.9.1 dataclasses-0.6 flatbuffers-2.0.7 flax-0.5.3 jax-0.3.1
WARNING: The following packages were previously imported in this runtime:
[dataclasses]
You must restart the runtime in order to use newly installed versions.

```

RESTART SESSION



```
!pip install brax==0.1.1 #install compatible brax version
```

```
# Install other dependencies
```

```
!pip install gym==0.23.1
```

```
!pip install pytorch-lightning
```

```
!pip install pyvirtualdisplay
```

```
!pip install protobuf==3.20.3
```

```
Looking in links: https://storage.googleapis.com/jax-releases/jax\_cuda\_releases.html
Requirement already satisfied: jax[cuda11_pip] in /usr/local/lib/python3.10/dist-packages (0.3.14)
Collecting jax[cuda11_pip]
  Downloading jax-0.4.37-py3-none-any.whl.metadata (22 kB)
WARNING: jax 0.4.37 does not provide the extra 'cuda11-pip'
Collecting jaxlib<=0.4.37,>=0.4.36 (from jax[cuda11_pip])
  Downloading jaxlib-0.4.36-cp310-cp310-manylinux2014_x86_64.whl.metadata (1.0 kB)
Requirement already satisfied: ml_dtypes>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from jax[cuda11_pip]) (0.4.1)
Requirement already satisfied: numpy>=1.24 in /usr/local/lib/python3.10/dist-packages (from jax[cuda11_pip]) (1.26.4)
Requirement already satisfied: opt_einsum in /usr/local/lib/python3.10/dist-packages (from jax[cuda11_pip]) (3.4.0)
Requirement already satisfied: scipy>=1.10 in /usr/local/lib/python3.10/dist-packages (from jax[cuda11_pip]) (1.13.1)
Downloading jaxlib-0.4.36-cp310-cp310-manylinux2014_x86_64.whl (100.3 MB)
100.3/100.3 MB 8.0 MB/s eta 0:00:00
Downloading jax-0.4.37-py3-none-any.whl (2.2 MB)
2.2/2.2 MB 52.1 MB/s eta 0:00:00
Installing collected packages: jaxlib, jax
  Attempting uninstall: jaxlib
    Found existing installation: jaxlib 0.3.14+cuda11.cudnn82
    Uninstalling jaxlib-0.3.14+cuda11.cudnn82:
      Successfully uninstalled jaxlib-0.3.14+cuda11.cudnn82
  Attempting uninstall: jax
    Found existing installation: jax 0.3.14
    Uninstalling jax-0.3.14:
      Successfully uninstalled jax-0.3.14
Successfully installed jax-0.4.37 jaxlib-0.4.36
Collecting brax==0.1.1
  Downloading brax-0.1.1-py3-none-any.whl.metadata (6.7 kB)
Requirement already satisfied: absl-py in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (1.4.0)
Requirement already satisfied: dataclasses in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (0.6)
Collecting dm-env (from brax==0.1.1)
  Downloading dm_env-1.6-py3-none-any.whl.metadata (966 bytes)
Requirement already satisfied: flax in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (0.5.3)
Requirement already satisfied: grpcio in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (1.68.1)
Requirement already satisfied: gym in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (0.23.1)
Requirement already satisfied: jax in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (0.4.37)
Requirement already satisfied: jaxlib in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (0.4.36)
Collecting jaxopt (from brax==0.1.1)
  Downloading jaxopt-0.8.3-py3-none-any.whl.metadata (2.6 kB)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (3.1.4)
Collecting mujoco (from brax==0.1.1)
  Downloading mujoco-3.2.6-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (44 kB)
44.4/44.4 kB 3.3 MB/s eta 0:00:00
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (1.26.4)
Requirement already satisfied: optax in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (0.1.7)
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (11.0.0)
Requirement already satisfied: pytinyrenderer in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (0.0.14)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (1.13.1)
Requirement already satisfied: tensorboardX in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (2.6.2.2)
Collecting trimesh==3.9.35 (from brax==0.1.1)
  Downloading trimesh-3.9.35-py3-none-any.whl.metadata (16 kB)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from brax==0.1.1) (4.12.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from trimesh==3.9.35->brax==0.1.1) (75.1.0)
Requirement already satisfied: dm-tree in /usr/local/lib/python3.10/dist-packages (from dm-env->brax==0.1.1) (0.1.8)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.1.1) (3.8.0)
Requirement already satisfied: msgpack in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.1.1) (1.1.0)
Requirement already satisfied: rich~=11.1 in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.1.1) (11.2.0)
Requirement already satisfied: tensorstore in /usr/local/lib/python3.10/dist-packages (from flax->brax==0.1.1) (0.1.69)
```

```
!pip install jax==0.3.14 jaxlib==0.3.14+cuda11.cudnn82 -f https://storage.googleapis.com/jax-releases/jax_cuda_releases.html flax==0.4.2
```

```
Looking in links: https://storage.googleapis.com/jax-releases/jax\_cuda\_releases.html
Collecting jax==0.3.14
  Using cached jax-0.3.14-py3-none-any.whl
Collecting jaxlib==0.3.14+cuda11.cudnn82
  Using cached https://storage.googleapis.com/jax-releases/cuda11/jaxlib-0.3.14%2Bcuda11.cudnn82-cp310-none-manylinux2014\_x86\_64.whl
Collecting flax==0.4.2
  Downloading flax-0.4.2-py3-none-any.whl.metadata (9.3 kB)
Requirement already satisfied: absl-py in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (1.4.0)
Requirement already satisfied: numpy>=1.19 in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (1.26.4)
Requirement already satisfied: opt_einsum in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (3.4.0)
Requirement already satisfied: scipy>=1.5 in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (1.14.1)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (4.12.2)
Requirement already satisfied: etils[epath] in /usr/local/lib/python3.10/dist-packages (from jax==0.3.14) (1.11.0)
Requirement already satisfied: flatbuffers<3.0,>=1.12 in /usr/local/lib/python3.10/dist-packages (from jaxlib==0.3.14+cuda11.cudnn82) (2.0.9)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from flax==0.4.2) (3.8.0)
Requirement already satisfied: msgpack in /usr/local/lib/python3.10/dist-packages (from flax==0.4.2) (1.1.0)
Requirement already satisfied: optax in /usr/local/lib/python3.10/dist-packages (from flax==0.4.2) (0.1.7)
Requirement already satisfied: fspec in /usr/local/lib/python3.10/dist-packages (from etils[epath]->jax==0.3.14) (2024.10.0)
Requirement already satisfied: importlib_resources in /usr/local/lib/python3.10/dist-packages (from etils[epath]->jax==0.3.14) (6.4
```



```
Requirement already satisfied: zipp in /usr/local/lib/python3.10/dist-packages (from etils[epath]->jax==0.3.14) (3.21.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (1.3.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (4.55.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (1.4.7)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (24.2)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->flax==0.4.2) (2.8.1)
Requirement already satisfied: chex>=0.1.5 in /usr/local/lib/python3.10/dist-packages (from optax->flax==0.4.2) (0.1.6)
Requirement already satisfied: dm-tree>=0.1.5 in /usr/local/lib/python3.10/dist-packages (from chex>=0.1.5->optax->flax==0.4.2) (0.1.8)
Requirement already satisfied: toolz>=0.9.0 in /usr/local/lib/python3.10/dist-packages (from chex>=0.1.5->optax->flax==0.4.2) (0.12.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib->flax==0.4.2) (1.16.0)
Downloading flax-0.4.2-py3-none-any.whl (186 kB)
```

186.4/186.4 kB 6.9 MB/s eta 0:00:00

Installing collected packages: jaxlib, jax, flax

Attempting uninstall: jaxlib

Found existing installation: jaxlib 0.4.36

Uninstalling jaxlib-0.4.36:

Successfully uninstalled jaxlib-0.4.36

Attempting uninstall: jax

Found existing installation: jax 0.4.37

Uninstalling jax-0.4.37:

Successfully uninstalled jax-0.4.37

Attempting uninstall: flax

Found existing installation: flax 0.6.9

Uninstalling flax-0.6.9:

Successfully uninstalled flax-0.6.9

ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency issues. orbax-checkpoint 0.6.4 requires jax>=0.4.26, but you have jax 0.3.14 which is incompatible.

Successfully installed flax-0.4.2 jax-0.3.14 jaxlib-0.3.14+cuda11.cudnn82

```
!pip install scipy==1.9.0
```



Collecting scipy==1.9.0

Downloading scipy-1.9.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (2.2 kB)

Collecting numpy<1.25.0,>=1.18.5 (from scipy==1.9.0)

Downloading numpy-1.24.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.6 kB)

Downloading scipy-1.9.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (43.9 MB)

43.9/43.9 MB 12.9 MB/s eta 0:00:00

Downloading numpy-1.24.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (17.3 MB)

17.3/17.3 MB 70.0 MB/s eta 0:00:00

Installing collected packages: numpy, scipy

Attempting uninstall: numpy

Found existing installation: numpy 1.26.4

Uninstalling numpy-1.26.4:

Successfully uninstalled numpy-1.26.4

Attempting uninstall: scipy

Found existing installation: scipy 1.14.1

Uninstalling scipy-1.14.1:

Successfully uninstalled scipy-1.14.1

ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency issues. albumentations 1.4.20 requires scipy>=1.10.0, but you have scipy 1.9.0 which is incompatible.

ibis-framework 9.2.0 requires rich<14,>=12.4.4, but you have rich 11.2.0 which is incompatible.

orbax-checkpoint 0.6.4 requires jax>=0.4.26, but you have jax 0.3.14 which is incompatible.

pymc 5.18.2 requires rich>=13.7.1, but you have rich 11.2.0 which is incompatible.

tensorflow 2.17.1 requires flatbuffers>=24.3.25, but you have flatbuffers 2.0.7 which is incompatible.

Successfully installed numpy-1.24.4 scipy-1.9.0

WARNING: The following packages were previously imported in this runtime:

[numpy]

You must restart the runtime in order to use newly installed versions.

RESTART SESSION

```
import warnings
warnings.filterwarnings('ignore')
```

Setup virtual display

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



```
<pyvirtualdisplay.display.Display at 0x793de1c67640>
```

Import the necessary code libraries

```
import copy
import torch
import random
```

```

import gym
import matplotlib
import functools
import math

import numpy as np
import matplotlib.pyplot as plt

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 torch.distributions import Normal


from pytorch_lightning import LightningModule, Trainer

import brax
from brax import envs
from brax.envs import to_torch
from brax.io import html

device = 'cuda:0' if torch.cuda.is_available() else 'cpu'
num_gpus = torch.cuda.device_count()

v = torch.ones(1, device='cpu')

```

 /usr/local/lib/python3.10/dist-packages/trimesh/interfaces/scad.py:28: DeprecationWarning: Use shutil.which instead of find_executable
 _scad_executable = find_executable('openscad', path=_search_path)
 /usr/local/lib/python3.10/dist-packages/trimesh/interfaces/blender.py:36: DeprecationWarning: Use shutil.which instead of find_executable
 _blender_executable = find_executable('blender', path=_search_path)
 /usr/local/lib/python3.10/dist-packages/trimesh/interfaces/vhacd.py:21: DeprecationWarning: Use shutil.which instead of find_executable
 _vhacd_executable = find_executable(_name, path=_search_path)
 /usr/local/lib/python3.10/dist-packages/trimesh/curvature.py:12: DeprecationWarning: Please use `coo_matrix` from the `scipy.sparse`
 from scipy.sparse.coo import coo_matrix
 /usr/local/lib/python3.10/dist-packages/trimesh/exchange/ply.py:953: DeprecationWarning: Use shutil.which instead of find_executable
 draco_encoder = find_executable('draco_encoder')
 /usr/local/lib/python3.10/dist-packages/trimesh/exchange/binvox.py:18: DeprecationWarning: Use shutil.which instead of find_executable
 binvox_encoder = find_executable('binvox')
 /usr/local/lib/python3.10/dist-packages/trimesh/path/exchange/dxf.py:1101: DeprecationWarning: Use shutil.which instead of find_executable
 _teigha = find_executable(_name)

```

@torch.no_grad()
def plot_policy(policy):
    pos = np.linspace(-4.8, 4.8, 100)
    vel = np.random.random(size=(10000, 1)) * 0.1
    ang = np.linspace(-0.418, 0.418, 100)
    ang_vel = np.random.random(size=(10000, 1)) * 0.1

    g1, g2 = np.meshgrid(pos, ang)
    grid = np.stack((g1, g2), axis=-1)
    grid = grid.reshape(-1, 2)
    grid = np.hstack((grid, vel, ang_vel))

    grid = torch.from_numpy(grid).float()
    loc, _ = policy(grid)

    plot_vals = loc.numpy()
    plot_vals = plot_vals.reshape(100, 100)[::-1]

    plt.figure(figsize=(8, 8))
    plt.imshow(plot_vals, cmap='coolwarm')
    plt.colorbar()
    plt.clim(-1, 1)
    plt.title("P(left | s)", size=20)
    plt.xlabel("Cart Position", size=14)
    plt.ylabel("Pole angle", size=14)
    plt.xticks(ticks=[0, 50, 100], labels=['-4.8', '0', '4.8'])
    plt.yticks(ticks=[100, 50, 0], labels=['-0.418', '0', '0.418'])

@torch.no_grad()
def create_video(env, episode_length, policy=None):
    qp_array = []
    state = env.reset()
    for i in range(episode_length):

```



```

    if policy:
        loc, scale = policy(state)
        sample = torch.normal(loc, scale)
        action = torch.tanh(sample)
    else:
        action = env.action_space.sample()
    state, _, _, _ = env.step(action)
    qp_array.append(env.unwrapped._state.qp)
return HTML(html.render(env.unwrapped._env.sys, qp_array))

@torch.no_grad()
def test_agent(env, episode_length, policy, episodes=10):
    ep_returns = []
    for ep in range(episodes):
        state = env.reset()
        done = False
        ep_ret = 0.0

        while not done:
            loc, scale = policy(state)
            sample = torch.normal(loc, scale)
            action = torch.tanh(sample)
            state, reward, done, info = env.step(action)
            ep_ret += reward.item()

        ep_returns.append(ep_ret)

    return sum(ep_returns) / episodes

```

✓ Create the policy

```

class GradientPolicy(nn.Module):

    def __init__(self, in_features, out_dims, hidden_size=128):
        super().__init__()
        self.fc1 = nn.Linear(in_features, hidden_size)
        self.fc2 = nn.Linear(hidden_size, hidden_size)
        self.fc_mu = nn.Linear(hidden_size, out_dims)
        self.fc_std = nn.Linear(hidden_size, out_dims)

    def forward(self, x):
        x = F.relu(self.fc1(x))
        x = F.relu(self.fc2(x))
        loc = self.fc_mu(x)
        loc = torch.tanh(loc)
        scale = self.fc_std(x)
        scale = F.softplus(scale) + 0.001
        return loc, scale

```

✓ Create the environment

```

class RunningMeanStd:
    # https://en.wikipedia.org/wiki/Algorithms_for_calculating_variance#Parallel_algorithm
    def __init__(self, epsilon=1e-4, shape=()):
        self.mean = torch.zeros(shape, dtype=torch.float32).to(device)
        self.var = torch.ones(shape, dtype=torch.float32).to(device)
        self.count = epsilon

    def update(self, x):
        batch_mean = torch.mean(x, dim=0)
        batch_var = torch.var(x, dim=0)
        batch_count = x.shape[0]
        self.update_from_moments(batch_mean, batch_var, batch_count)

    def update_from_moments(self, batch_mean, batch_var, batch_count):
        self.mean, self.var, self.count = update_mean_var_count_from_moments(
            self.mean, self.var, self.count, batch_mean, batch_var, batch_count
        )

def update_mean_var_count_from_moments(
    mean, var, count, batch_mean, batch_var, batch_count
):
    delta = batch_mean - mean
    tot_count = count + batch_count

    new_mean = mean + delta * batch_count / tot_count

```

```

m_a = var * count
m_b = batch_var * batch_count
M2 = m_a + m_b + torch.square(delta) * count * batch_count / tot_count
new_var = M2 / tot_count
new_count = tot_count

return new_mean, new_var, new_count

class NormalizeObservation(gym.core.Wrapper):

    def __init__(self, env, epsilon=1e-8):
        super().__init__(env)
        self.num_envs = getattr(env, "num_envs", 1)
        self.obs_rms = RunningMeanStd(shape=self.observation_space.shape[-1])
        self.epsilon = epsilon

    def step(self, action):
        obs, rews, dones, infos = self.env.step(action)
        obs = self.normalize(obs)
        return obs, rews, dones, infos

    def reset(self, **kwargs):
        return_info = kwargs.get("return_info", False)
        if return_info:
            obs, info = self.env.reset(**kwargs)
        else:
            obs = self.env.reset(**kwargs)
        obs = self.normalize(obs)
        if not return_info:
            return obs
        else:
            return obs, info

    def normalize(self, obs):
        self.obs_rms.update(obs)
        return (obs - self.obs_rms.mean) / torch.sqrt(self.obs_rms.var + self.epsilon)

class NormalizeReward(gym.core.Wrapper):

    def __init__(self, env, gamma=0.99, epsilon=1e-8):
        super().__init__(env)
        self.num_envs = getattr(env, "num_envs", 1)
        self.return_rms = RunningMeanStd(shape=())
        self.returns = torch.zeros(self.num_envs).to(device)
        self.gamma = gamma
        self.epsilon = epsilon

    def step(self, action):
        obs, rews, dones, infos = self.env.step(action)
        dones = dones.bool()
        self.returns = self.returns * self.gamma + rews
        rews = self.normalize(rews)
        self.returns[dones] = 0.0
        return obs, rews, dones, infos

    def normalize(self, rews):
        self.return_rms.update(self.returns)
        return rews / torch.sqrt(self.return_rms.var + self.epsilon)

```

```

entry_point = functools.partial(envs.create_gym_env, env_name='inverted_pendulum')
gym.register('brax-inverted_pendulum-v0', entry_point=entry_point)

```

```

def create_env(env_name, num_envs=256, episode_length=1000):
    env = gym.make(env_name, batch_size=num_envs, episode_length=episode_length)
    env = to_torch.JaxToTorchWrapper(env, device=device)
    env = NormalizeObservation(env)
    env = NormalizeReward(env)
    return env

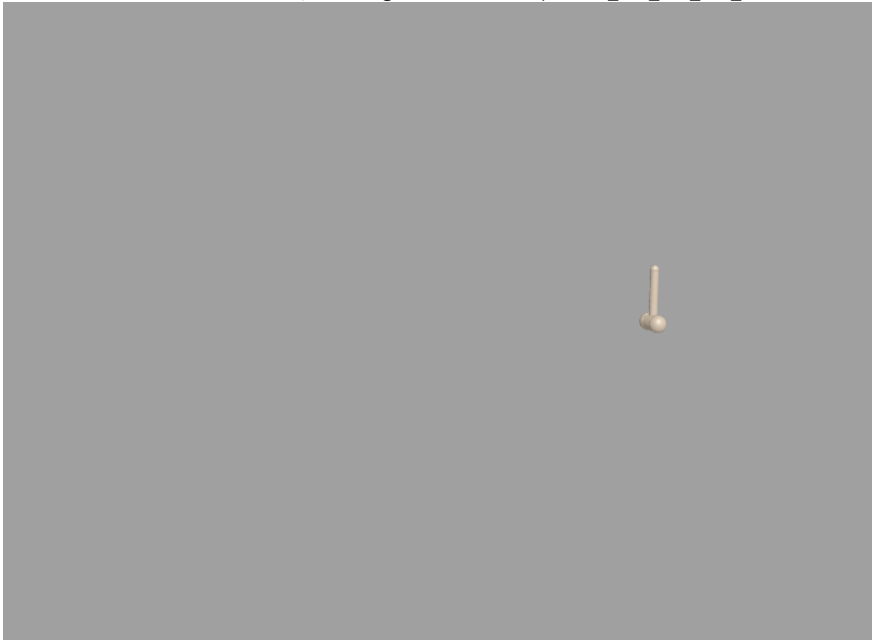
```

```

env = gym.make('brax-inverted_pendulum-v0', episode_length=1000)
env = to_torch.JaxToTorchWrapper(env, device=device)
create_video(env, 1000)

```

WARNING:absl:No GPU/TPU found, falling back to CPU. (Set TF_CPP_MIN_LOG_LEVEL=0 and rerun for more info) > Controls



```
env = create_env('brax-inverted_pendulum-v0', num_envs=1)
obs = env.reset()
print("Num envs: ", obs.shape[0], "Obs dimentions: ", obs.shape[1])
```

Num envs: 1 Obs dimentions: 4

```
env.action_space
```

Box(-1.0, 1.0, (1, 1), float32)

```
obs, reward, done, info = env.step(env.action_space.sample())
```

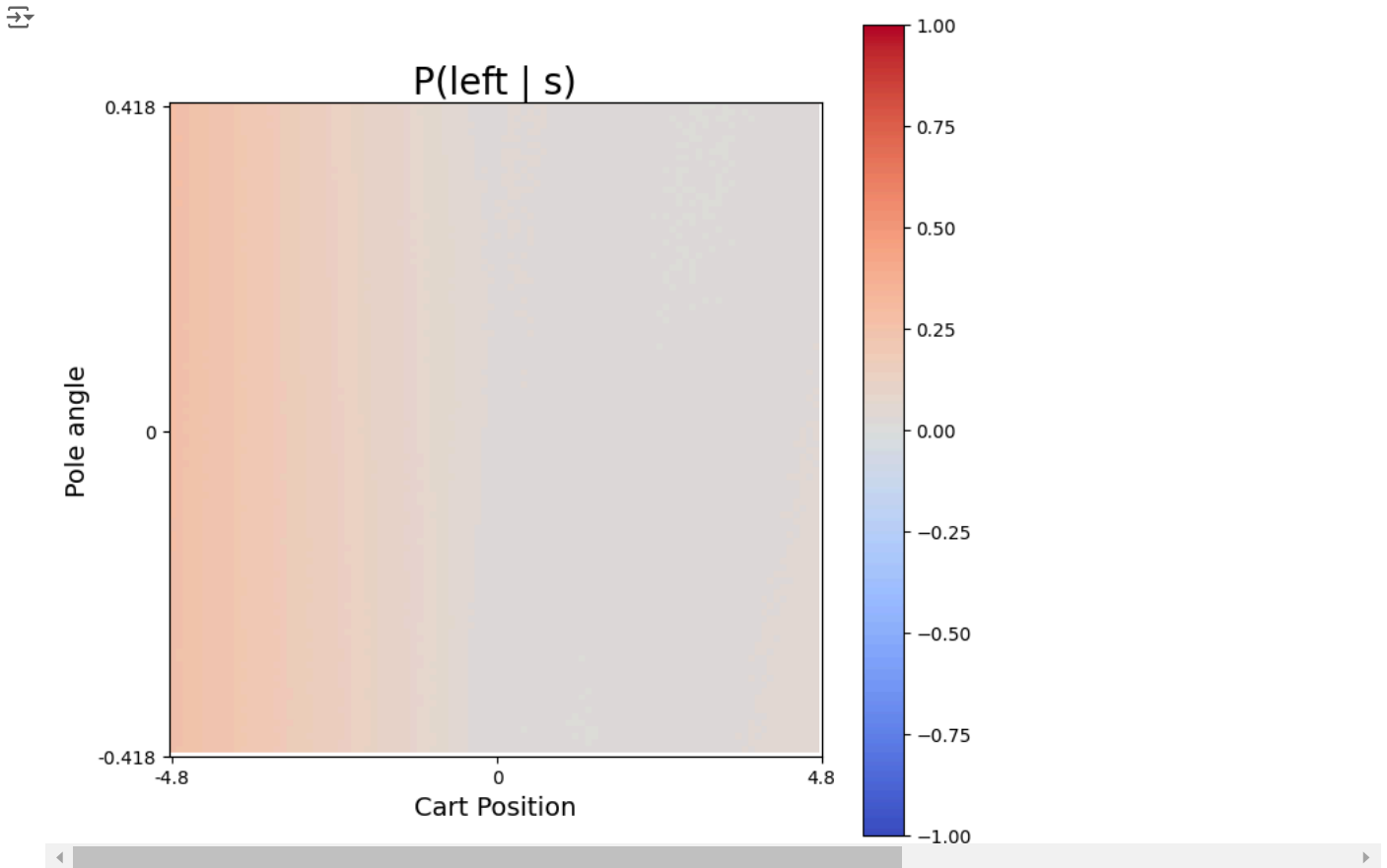
```
info.keys()
```

dict_keys(['first_obs', 'first_qp', 'steps', 'truncation'])

Start coding or [generate](#) with AI.

Plot the untrained policy

```
policy = GradientPolicy(4, 1)
grid = plot_policy(policy)
```



▼ Create the dataset

```
class RLDataset(IterableDataset):

    def __init__(self, env, policy, episode_length, gamma):
        self.env = env
        self.policy = policy
        self.episode_length = episode_length
        self.gamma = gamma
        self.obs = self.env.reset()

    @torch.no_grad()
    def __iter__(self):
        transitions = []

        for step in range(self.episode_length):
            loc, scale = self.policy(self.obs)
            action = torch.normal(loc, scale)
            next_obs, reward, done, info = self.env.step(action)
            transitions.append((self.obs, action, reward, done))
            self.obs = next_obs

        obs_b, action_b, reward_b, done_b = map(torch.stack, zip(*transitions))

        running_return = torch.zeros(self.env.num_envs, dtype=torch.float32, device=device)
        return_b = torch.zeros_like(reward_b)

        for row in range(self.episode_length - 1, -1, -1):
            running_return = reward_b[row] + ~done_b[row] * self.gamma * running_return
            return_b[row] = running_return

        num_samples = self.env.num_envs * self.episode_length
        obs_b = obs_b.view(num_samples, -1)
        action_b = action_b.view(num_samples, -1)
        return_b = return_b.view(num_samples, -1)

        idx = list(range(num_samples))
        random.shuffle(idx)

        for i in idx:
            yield obs_b[i], action_b[i], return_b[i]
```

▼ Create the REINFORCE algorithm

```

class reinforce(LightningModule):

    def __init__(self, env_name, num_envs=256, episode_length=1_000, batch_size=1024,
                 hidden_size=64, policy_lr=1e-4, gamma=0.999, entropy_coef=0.0001, optim=AdamW):

        super().__init__()

        self.env = create_env(env_name, num_envs=num_envs, episode_length=episode_length)
        test_env = gym.make(env_name, episode_length=episode_length)
        test_env = to_torch.JaxToTorchWrapper(test_env, device=device)
        self.test_env = NormalizeObservation(test_env)
        self.test_env.obs_rms = self.env.obs_rms

        obs_size = self.env.observation_space.shape[1]
        action_dims = self.env.action_space.shape[1]

        self.policy = GradientPolicy(obs_size, action_dims, hidden_size)
        self.dataset = RLDataset(self.env, self.policy, episode_length, gamma)

        self.save_hyperparameters()
        self.videos = []

    def configure_optimizers(self):
        return self.hparams.optim(self.policy.parameters(), lr=self.hparams.policy_lr)

    def train_dataloader(self):
        return DataLoader(dataset=self.dataset, batch_size=self.hparams.batch_size)

    # Training step.
    def training_step(self, batch, batch_idx):
        obs, action, returns = batch

        loc, scale = self.policy(obs)
        dist = Normal(loc, scale)
        log_prob = dist.log_prob(action).sum(dim=-1, keepdim=True)

        policy_loss = - log_prob * returns

        entropy = dist.entropy().sum(dim=-1, keepdim=True)

        self.log("episode/Policy Loss", policy_loss.mean())
        self.log("episode/Entropy", entropy.mean())

        return torch.mean(policy_loss - self.hparams.entropy_coef * entropy)

    def on_train_epoch_end(self):
        if self.current_epoch % 10 == 0:
            average_return = test_agent(self.test_env, self.hparams.episode_length, self.policy, episodes=1)
            self.log("episode/Average Return", average_return)

        if self.current_epoch % 50 == 0:
            video = create_video(self.test_env, self.hparams.episode_length, policy=self.policy)
            self.videos.append(video)

```