In [2]:
```
%load_ext autoreload
%autoreload 2
```

In [3]:
```
import gym
import torch
import numpy as np
from loguru import logger
from torch.optim import Adam
import matplotlib.pyplot as plt
from IPython.display import Image

from src.utils import (
    device,
    set_seed,
    eval_policy,
    demo_policy,
    plot_returns,
    save_frames_as_gif
)

plt.ion()
```

Out[3]:
```
<contextlib.ExitStack at 0x7211484d93c0>
```

In [8]:
```
SEED: int = 42
ENVIRONMENT_NAME: str='LunarLander-v2'

# torch related defaults
DEVICE = device()
torch.set_default_dtype(torch.float32)
```

```
2024-10-27 20:52:32.691 | INFO     | src.utils:device:51 - Using cpu device.
```

In [13]:
```
# Use random seeds for reproducibility
set_seed(SEED)

# instantiate the environment
environment = gym.make(ENVIRONMENT_NAME)

# get the state and action dimensions
num_actions = environment.action_space.n
state_dimension = environment.observation_space.shape[0]
```

```
2024-10-27 20:52:47.209 | INFO     | src.utils:set_seed:37 - Random seed set as 42.
```

---

1. REINFORCE

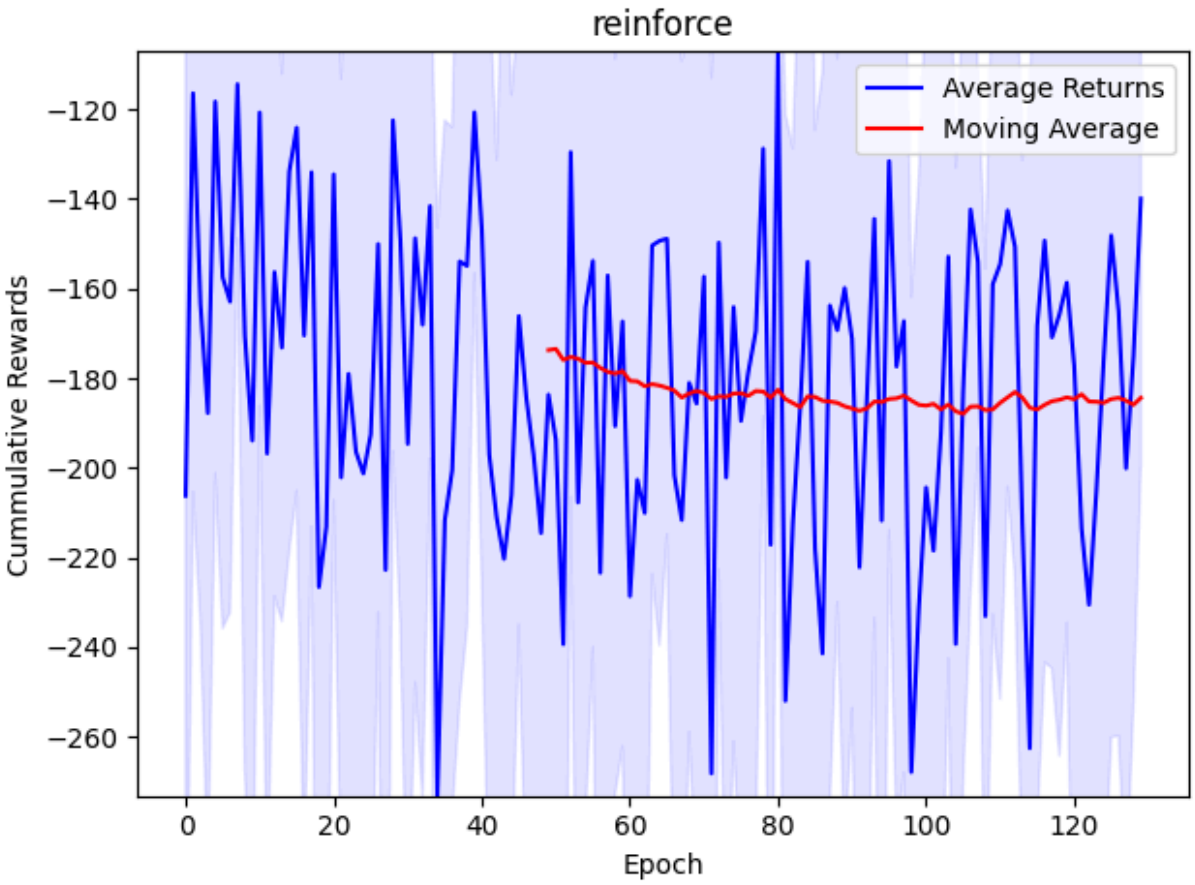
In [59]:
```
from src.networks import Policy
from src.reinforce import train_one_epoch as reinforce_epoch
```

```
################################## Hyper-parameters Tuning ################

EPOCHS: int = 1000
HIDDEN_DIMENSION: int = 64
LEARNING_RATE: float = 1e-3

###########################################################################

# Instantiate the policy network
policy = Policy(
    state_dimension, num_actions, hidden_dimension=HIDDEN_DIMENSION
).to(DEVICE)

# Learn the policy
optimizer = Adam(policy.parameters(), LEARNING_RATE)

# Loop for each epoch
mean_returns, std_returns = [], []
for epoch in range(EPOCHS):
    reinforce_epoch(environment, policy, optimizer)

    episode_return_mean, episode_return_std = eval_policy(policy, ENVIRONMEN
    mean_returns.append(episode_return_mean)
    std_returns.append(episode_return_std)

    logger.info(f'Epoch: {epoch:3d} \t return: {episode_return_mean:.2f}')

    if epoch:
        plot_returns(
            mean_returns, std_returns, method_name='reinforce', dynamic=True
        )
```

```
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
Cell In[59], line 23
     21 mean_returns, std_returns = [], []
     22 for epoch in range(EPOCHS):
---> 23     reinforce_epoch(environment, policy, optimizer)
     25     episode_return_mean, episode_return_std = eval_policy(policy, EN
VIRONMENT_NAME)
     26     mean_returns.append(episode_return_mean)

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/reinforce.p
y:57, in train_one_epoch(env, policy, optimizer, max_timesteps)
     54 epoch_total_timesteps += 1
     56 # TODO: Sample an action from the policy
---> 57 action, log_prob_of_action = policy.sample(tensor(state))
     58 # print('action', action)
     59 # TODO: Take the action in the environment
     60 state, reward, terminated, truncated, info = env.step(action)

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/networks.py:
105, in Policy.sample(self, state)
     93 def sample(self, state: np.ndarray) -> Tuple[int, torch.Tensor]:
     94     """
     95     Samples an action from the policy and returns the action along w
ith its log probability.
     96
   (...)
    103     TODO: Implement the sample method to sample an action and comput
e its log probability.
    104     """
--> 105     PI = self.pi(state)
    106     action = PI.sample()
    107     log_prob_of_action = PI.log_prob(action)

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/networks.py:
89, in Policy.pi(self, state)
     77 def pi(self, state: np.ndarray) -> Categorical:
     78     """
     79     Computes the action distribution π(a|s) for a given state.
     80
   (...)
     87     TODO: Implement the pi method to create a Categorical distributi
on based on the network's output.
     88     """
---> 89     action_logits = self.forward(state)
     90     return Categorical(logits = action_logits)

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/networks.py:
74, in Policy.forward(self, state)
     61 """
     62 Forward pass of the Policy network.
     63
   (...)
     71 You can use the self.network to forward the input.
     72 """
     73 state = tensor(state)
```

```
---> 74 return self.network(state)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/torch/nn/modules/module.py:1532, in Module._wrapped_call_impl(self, *args, **kwargs)
   1530     return self._compiled_call_impl(*args, **kwargs)  # type: ignore[misc]
   1531 else:
-> 1532     return self._call_impl(*args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/torch/nn/modules/module.py:1541, in Module._call_impl(self, *args, **kwargs)
   1536 # If we don't have any hooks, we want to skip the rest of the logic in
   1537 # this function, and just call forward.
   1538 if not (self._backward_hooks or self._backward_pre_hooks or self._forward_hooks or self._forward_pre_hooks
   1539         or _global_backward_pre_hooks or _global_backward_hooks
   1540         or _global_forward_hooks or _global_forward_pre_hooks):
-> 1541     return forward_call(*args, **kwargs)
   1543 try:
   1544     result = None

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/networks.py:39, in network.<locals>.ShallowNet.forward(self, x)
     38 def forward(self, x):
---> 39     x = F.relu(self.fc1(x))
     40     logits = self.fc2(x)
     41     return logits

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/torch/nn/modules/module.py:1532, in Module._wrapped_call_impl(self, *args, **kwargs)
   1530     return self._compiled_call_impl(*args, **kwargs)  # type: ignore[misc]
   1531 else:
-> 1532     return self._call_impl(*args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/torch/nn/modules/module.py:1541, in Module._call_impl(self, *args, **kwargs)
   1536 # If we don't have any hooks, we want to skip the rest of the logic in
   1537 # this function, and just call forward.
   1538 if not (self._backward_hooks or self._backward_pre_hooks or self._forward_hooks or self._forward_pre_hooks
   1539         or _global_backward_pre_hooks or _global_backward_hooks
   1540         or _global_forward_hooks or _global_forward_pre_hooks):
-> 1541     return forward_call(*args, **kwargs)
   1543 try:
   1544     result = None

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/torch/nn/modules/linear.py:116, in Linear.forward(self, input)
    115 def forward(self, input: Tensor) -> Tensor:
--> 116     return F.linear(input, self.weight, self.bias)

KeyboardInterrupt:
<Figure size 640x480 with 0 Axes>
```

In [60]:
```
plot_returns(
    mean_returns, std_returns, method_name='reinforce'
)
```

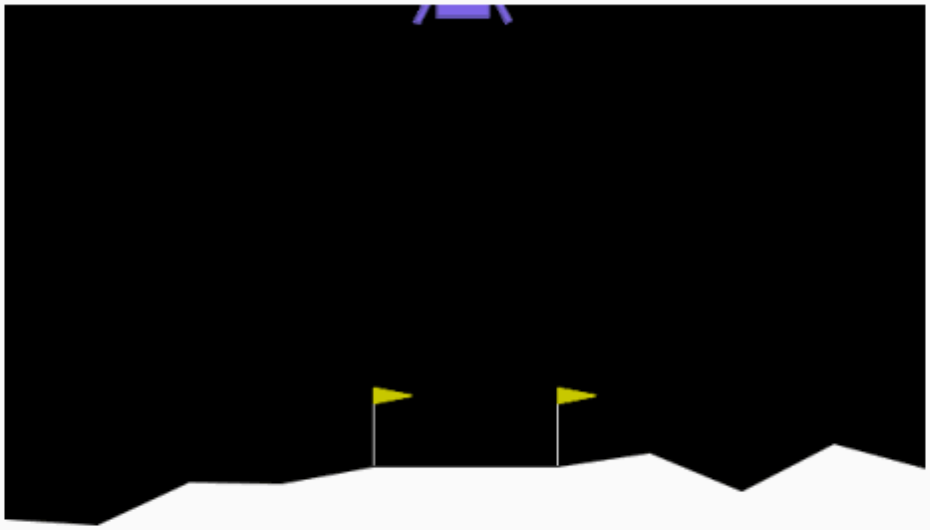



```
<Figure size 640x480 with 0 Axes>
<Figure size 640x480 with 0 Axes>
```

In [61]:
```
frames = demo_policy(
    policy, ENVIRONMENT_NAME
)
gif_path = save_frames_as_gif(frames, method_name='reinforce')
Image(open(gif_path,'rb').read())
```

```
MovieWriter imagemagick unavailable; using Pillow instead.
```

Out[61]:



---

2. Simple Q-iteration (no experience replay + target network)

In [49]:

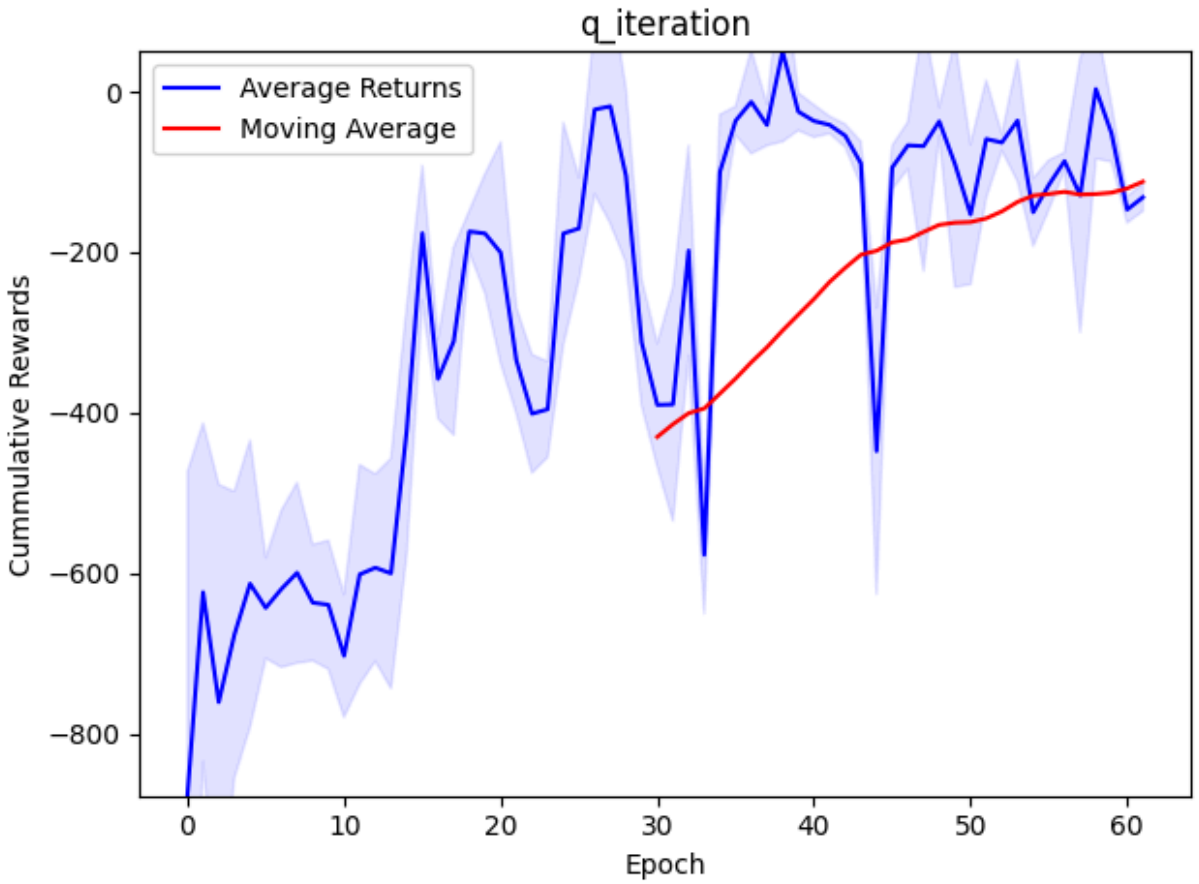
```
from src.networks import ValueFunctionQ
from src.q_iter import train_one_epoch as q_iter_epoch

################################### Hyper-parameters Tuning ################

EPOCHS: int = 1000
HIDDEN_DIMENSION: int = 64
LEARNING_RATE: float = 1e-3

############################################################################

# Instantiate the state-action value function, Q
Q = ValueFunctionQ(
    state_dimension, num_actions, hidden_dimension=HIDDEN_DIMENSION
).to(DEVICE)
# Learn the policy
optimizer = Adam(Q.parameters(), LEARNING_RATE)

mean_returns, std_returns = [], []
for epoch in range(EPOCHS):
    q_iter_epoch(SEED, env=environment, Q=Q, optimizer=optimizer)

    episode_return_mean, episode_return_std = eval_policy(Q, ENVIRONMENT_NAM
    mean_returns.append(episode_return_mean)
    std_returns.append(episode_return_std)
```

```
    from src.q_iter import eps

    logger.info(f'Epoch: {epoch:3d}, \t return: {episode_return_mean:.2f}, \

    if epoch:
        plot_returns(
            mean_returns, std_returns, method_name='q_iteration', dynamic=Tr
        )
```

```
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
Cell In[49], line 23
     20 for epoch in range(EPOCHS):
     21     q_iter_epoch(SEED, env=environment, Q=Q, optimizer=optimizer)
---> 23     episode_return_mean, episode_return_std = eval_policy(Q, ENVIRON
MENT_NAME)
     24     mean_returns.append(episode_return_mean)
     25     std_returns.append(episode_return_std)

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/utils.py:70,
in eval_policy(policy, environment_name)
     68 for step in range(MAX_EVAL_EPISODE_STEPS):
     69     action = policy.action(state)
---> 70     state, reward, terminated, truncated, info = eval_environment.st
ep(action)
     71     done = terminated or truncated
     72     episode_reward += reward

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/gym/wrapper
s/time_limit.py:50, in TimeLimit.step(self, action)
     39 def step(self, action):
     40     """Steps through the environment and if the number of steps elap
sed exceeds ``max_episode_steps`` then truncate.
     41
     42     Args:
   (...)
     48
     49     """
---> 50     observation, reward, terminated, truncated, info = self.env.step
(action)
     51     self._elapsed_steps += 1
     53     if self._elapsed_steps >= self._max_episode_steps:

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/gym/wrapper
s/order_enforcing.py:37, in OrderEnforcing.step(self, action)
     35 if not self._has_reset:
     36     raise ResetNeeded("Cannot call env.step() before calling env.res
et()")
---> 37 return self.env.step(action)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/gym/wrapper
s/env_checker.py:39, in PassiveEnvChecker.step(self, action)
     37     return env_step_passive_checker(self.env, action)
     38 else:
---> 39     return self.env.step(action)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/gym/envs/box
2d/lunar_lander.py:556, in LunarLander.step(self, action)
    545     p.ApplyLinearImpulse(
    546         (ox * SIDE_ENGINE_POWER * s_power, oy * SIDE_ENGINE_POWER *
s_power),
    547         impulse_pos,
    548         True,
    549     )
    550     self.lander.ApplyLinearImpulse(
```

```
    551         (-ox * SIDE_ENGINE_POWER * s_power, -oy * SIDE_ENGINE_POWER
* s_power),
    552         impulse_pos,
    553         True,
    554     )
--> 556 self.world.Step(1.0 / FPS, 6 * 30, 2 * 30)
    558 pos = self.lander.position
    559 vel = self.lander.linearVelocity

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/gym/envs/box
2d/lunar_lander.py:59, in ContactDetector.BeginContact(self, contact)
     56     contactListener.__init__(self)
     57     self.env = env
---> 59 def BeginContact(self, contact):
     60     if (
     61         self.env.lander == contact.fixtureA.body
     62         or self.env.lander == contact.fixtureB.body
     63     ):
     64         self.env.game_over = True

KeyboardInterrupt:
<Figure size 640x480 with 0 Axes>
```

In [50]:
```
plot_returns(
    mean_returns, std_returns, method_name='q_iteration'
)
```

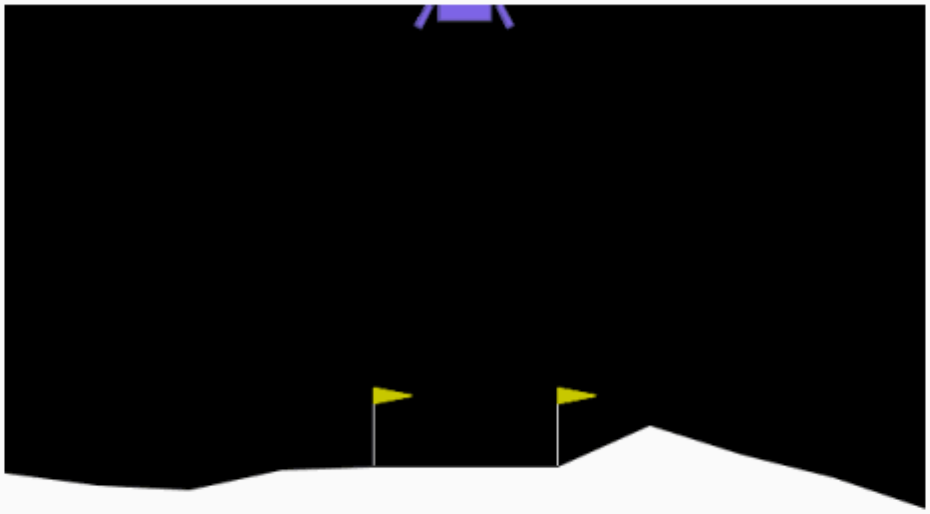



```
<Figure size 640x480 with 0 Axes>
<Figure size 640x480 with 0 Axes>
```

In [51]:
```
frames = demo_policy(
    Q, ENVIRONMENT_NAME
)
gif_path = save_frames_as_gif(frames, method_name='q_iteration')
Image(open(gif_path,'rb').read())
```

```
MovieWriter imagemagick unavailable; using Pillow instead.
```

Out[51]:



Experience Replay Buffer/Memory

In [52]:
```
from src.buffer import ReplayBuffer
################################### Hyper-parameters Tuning ################
BATCH_SIZE: int = 64
############################################################################

# instantiate the memory replay buffer
memory = ReplayBuffer(
    capacity=10_000, batch_size=BATCH_SIZE
)
```

---

3. DQN (Deep Q-learning + experience replay + target network)

In [53]:
```
from src.networks import ValueFunctionQ
from src.dqn import train_one_epoch as dqn_epoch

################################### Hyper-parameters Tuning ################

EPOCHS: int = 600
HIDDEN_DIMENSION: int = 64
LEARNING_RATE: float = 1e-3

############################################################################

# instantiate the state-action value function, Q
Q = ValueFunctionQ(
    state_dimension, num_actions, hidden_dimension=HIDDEN_DIMENSION
).to(DEVICE)
# initialize the target network
target_Q = ValueFunctionQ(
    state_dimension, num_actions, hidden_dimension=HIDDEN_DIMENSION
).to(DEVICE)

# Learn the policy
optimizer = Adam(Q.parameters(), LEARNING_RATE)

memory.clear()
mean_returns, std_returns = [], []
for epoch in range(EPOCHS):

    # copy target network params
    target_Q.load_state_dict(Q.state_dict())

    dqn_epoch(seed = SEED,
        env=environment, Q=Q, target_Q=target_Q,
        RB=memory, optimizer=optimizer
    )

    episode_return_mean, episode_return_std = eval_policy(Q, ENVIRONMENT_NAM
    mean_returns.append(episode_return_mean)
    std_returns.append(episode_return_std)

    from src.dqn import eps

    logger.info(f'Epoch: {epoch:3d}, \t return: {episode_return_mean:.2f}, \

    if epoch:
        plot_returns(
            mean_returns, std_returns, method_name='dqn', dynamic=True
        )
```

```
2024-10-27 23:16:56.167 | INFO     | __main__:<module>:42 - Epoch: 162,
return: -330.17,         eps: 0.19
<Figure size 640x480 with 0 Axes>
```

```
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
Cell In[53], line 45
     42 logger.info(f'Epoch: {epoch:3d}, \t return: {episode_return_mean:.2
f}, \t eps: {eps:.2f}')
     44 if epoch:
---> 45     plot_returns(
     46         mean_returns, std_returns, method_name='dqn', dynamic=True
     47     )

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/utils.py:12
4, in plot_returns(mean_returns, std_returns, method_name, dynamic)
    122 plt.title(method_name)
    123 plt.tight_layout()
--> 124 plt.show()
    126 plt.pause(0.01)  # pause a bit so that plots are updated
    127 display.display(plt.gcf())

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/p
yplot.py:612, in show(*args, **kwargs)
    568 """
    569 Display all open figures.
    570
   (...)
    609 explicitly there.
    610 """
    611 _warn_if_gui_out_of_main_thread()
--> 612 return _get_backend_mod().show(*args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib_i
nline/backend_inline.py:90, in show(close, block)
     88 try:
     89     for figure_manager in Gcf.get_all_fig_managers():
---> 90         display(
     91             figure_manager.canvas.figure,
     92             metadata=_fetch_figure_metadata(figure_manager.canvas.fi
gure)
     93         )
     94 finally:
     95     show._to_draw = []

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/display_functions.py:298, in display(include, exclude, metadata, transien
t, display_id, raw, clear, *objs, **kwargs)
    296     publish_display_data(data=obj, metadata=metadata, **kwargs)
    297 else:
--> 298     format_dict, md_dict = format(obj, include=include, exclude=excl
ude)
    299     if not format_dict:
    300         # nothing to display (e.g. _ipython_display_ took over)
    301         continue

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/formatters.py:182, in DisplayFormatter.format(self, obj, include, exclude)
    180 md = None
    181 try:
```

```
--> 182     data = formatter(obj)
    183 except:
    184     # FIXME: log the exception
    185     raise

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/decorator.p
y:232, in decorate.<locals>.fun(*args, **kw)
    230 if not kwsyntax:
    231     args, kw = fix(args, kw, sig)
--> 232 return caller(func, *(extras + args), **kw)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/formatters.py:226, in catch_format_error(method, self, *args, **kwargs)
    224 """show traceback on failed format call"""
    225 try:
--> 226     r = method(self, *args, **kwargs)
    227 except NotImplementedError:
    228     # don't warn on NotImplementedErrors
    229     return self._check_return(None, args[0])

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/formatters.py:343, in BaseFormatter.__call__(self, obj)
    341     pass
    342 else:
--> 343     return printer(obj)
    344 # Finally look for special method names
    345 method = get_real_method(obj, self.print_method)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/pylabtools.py:170, in print_figure(fig, fmt, bbox_inches, base64, **kwarg
s)
    167     from matplotlib.backend_bases import FigureCanvasBase
    168     FigureCanvasBase(fig)
--> 170 fig.canvas.print_figure(bytes_io, **kw)
    171 data = bytes_io.getvalue()
    172 if fmt == 'svg':

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/b
ackend_bases.py:2204, in FigureCanvasBase.print_figure(self, filename, dpi,
facecolor, edgecolor, orientation, format, bbox_inches, pad_inches, bbox_ext
ra_artists, backend, **kwargs)
   2200 try:
   2201     # _get_renderer may change the figure dpi (as vector formats
   2202     # force the figure dpi to 72), so we need to set it again here.
   2203     with cbook._setattr_cm(self.figure, dpi=dpi):
-> 2204         result = print_method(
   2205             filename,
   2206             facecolor=facecolor,
   2207             edgecolor=edgecolor,
   2208             orientation=orientation,
   2209             bbox_inches_restore=_bbox_inches_restore,
   2210             **kwargs)
   2211 finally:
   2212     if bbox_inches and restore_bbox:

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/b
```

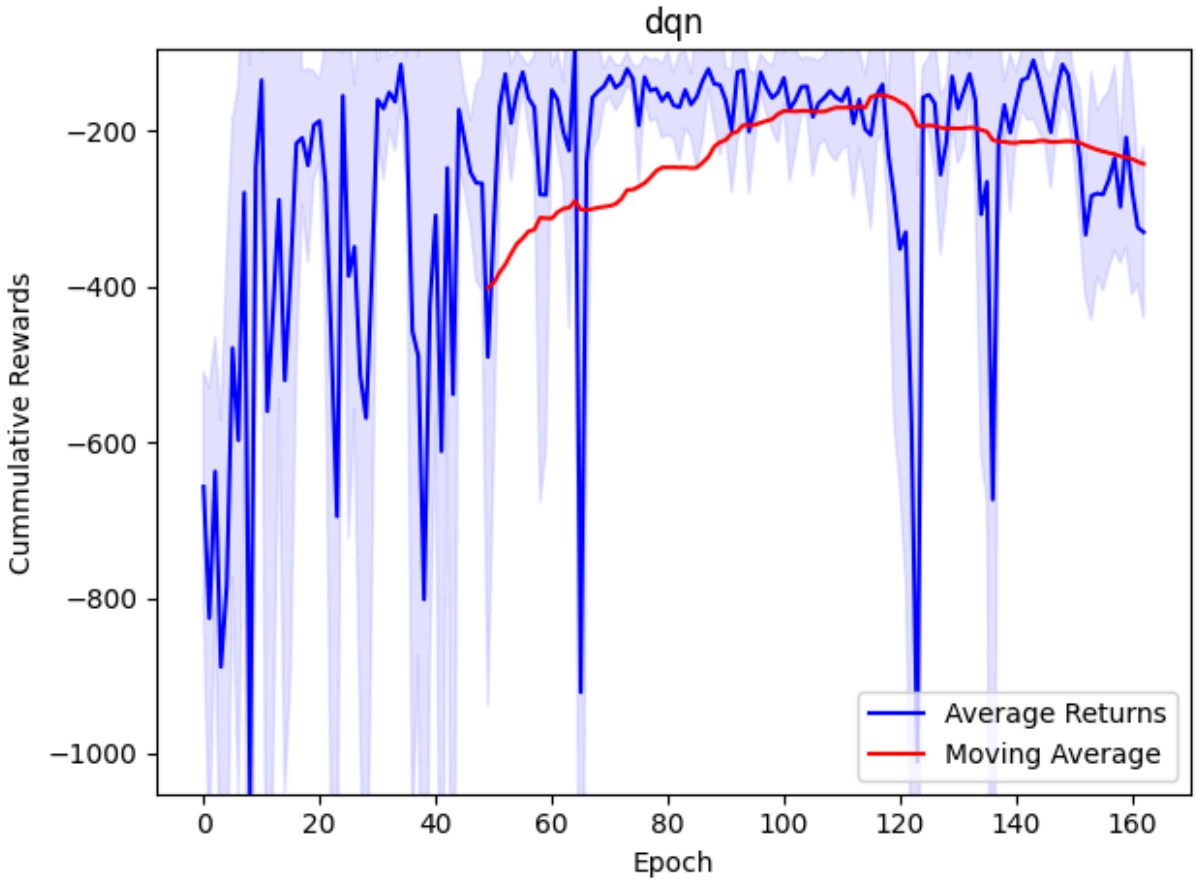
```
ackend_bases.py:2054, in FigureCanvasBase._switch_canvas_and_return_print_me
thod.<locals>.<lambda>(*args, **kwargs)
   2050     optional_kws = {  # Passed by print_figure for other renderers.
   2051         "dpi", "facecolor", "edgecolor", "orientation",
   2052         "bbox_inches_restore"}
   2053     skip = optional_kws - {*inspect.signature(meth).parameters}
-> 2054     print_method = functools.wraps(meth)(lambda *args, **kwargs: met
h(
   2055         *args, **{k: v for k, v in kwargs.items() if k not in ski
p}))
   2056 else:  # Let third-parties do as they see fit.
   2057     print_method = meth

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/b
ackends/backend_agg.py:496, in FigureCanvasAgg.print_png(self, filename_or_o
bj, metadata, pil_kwargs)
    449 def print_png(self, filename_or_obj, *, metadata=None, pil_kwargs=No
ne):
    450     """
    451     Write the figure to a PNG file.
    452
   (...)
    494         *metadata*, including the default 'Software' key.
    495     """
--> 496     self._print_pil(filename_or_obj, "png", pil_kwargs, metadata)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/b
ackends/backend_agg.py:444, in FigureCanvasAgg._print_pil(self, filename_or_
obj, fmt, pil_kwargs, metadata)
    439 def _print_pil(self, filename_or_obj, fmt, pil_kwargs, metadata=Non
e):
    440     """
    441     Draw the canvas, then save it using `.image.imsave` (to which
    442     *pil_kwargs* and *metadata* are forwarded).
    443     """
--> 444     FigureCanvasAgg.draw(self)
    445     mpl.image.imsave(
    446         filename_or_obj, self.buffer_rgba(), format=fmt, origin="upp
er",
    447         dpi=self.figure.dpi, metadata=metadata, pil_kwargs=pil_kwarg
s)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/b
ackends/backend_agg.py:387, in FigureCanvasAgg.draw(self)
    384 # Acquire a lock on the shared font cache.
    385 with (self.toolbar._wait_cursor_for_draw_cm() if self.toolbar
    386       else nullcontext()):
--> 387     self.figure.draw(self.renderer)
    388     # A GUI class may be need to update a window using this draw, so
    389     # don't forget to call the superclass.
    390     super().draw()

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
rtist.py:95, in _finalize_rasterization.<locals>.draw_wrapper(artist, render
er, *args, **kwargs)
     93 @wraps(draw)
```

```
     94 def draw_wrapper(artist, renderer, *args, **kwargs):
---> 95     result = draw(artist, renderer, *args, **kwargs)
     96     if renderer._rasterizing:
     97         renderer.stop_rasterizing()

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/artist.py:72, in allow_rasterization.<locals>.draw_wrapper(artist, renderer)
     69     if artist.get_agg_filter() is not None:
     70         renderer.start_filter()
---> 72     return draw(artist, renderer)
     73 finally:
     74     if artist.get_agg_filter() is not None:

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/figure.py:3162, in Figure.draw(self, renderer)
   3159             # ValueError can occur when resizing a window.
   3161     self.patch.draw(renderer)
-> 3162     mimage._draw_list_compositing_images(
   3163         renderer, self, artists, self.suppressComposite)
   3165     renderer.close_group('figure')
   3166 finally:

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/image.py:132, in _draw_list_compositing_images(renderer, parent, artists, suppress_composite)
    130 if not_composite or not has_images:
    131     for a in artists:
--> 132         a.draw(renderer)
    133 else:
    134     # Composite any adjacent images together
    135     image_group = []

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/artist.py:72, in allow_rasterization.<locals>.draw_wrapper(artist, renderer)
     69     if artist.get_agg_filter() is not None:
     70         renderer.start_filter()
---> 72     return draw(artist, renderer)
     73 finally:
     74     if artist.get_agg_filter() is not None:

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/axes/_base.py:3137, in _AxesBase.draw(self, renderer)
   3134 if artists_rasterized:
   3135     _draw_rasterized(self.figure, artists_rasterized, renderer)
-> 3137 mimage._draw_list_compositing_images(
   3138     renderer, self, artists, self.figure.suppressComposite)
   3140 renderer.close_group('axes')
   3141 self.stale = False

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/image.py:132, in _draw_list_compositing_images(renderer, parent, artists, suppress_composite)
    130 if not_composite or not has_images:
    131     for a in artists:
--> 132         a.draw(renderer)
    133 else:
```

```
    134     # Composite any adjacent images together
    135     image_group = []

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
rtist.py:72, in allow_rasterization.<locals>.draw_wrapper(artist, renderer)
     69     if artist.get_agg_filter() is not None:
     70         renderer.start_filter()
---> 72     return draw(artist, renderer)
     73 finally:
     74     if artist.get_agg_filter() is not None:

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/l
egend.py:777, in Legend.draw(self, renderer)
    774     Shadow(self.legendPatch, **self._shadow_props).draw(renderer)
    776 self.legendPatch.draw(renderer)
--> 777 self._legend_box.draw(renderer)
    779 renderer.close_group('legend')
    780 self.stale = False

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
rtist.py:39, in _prevent_rasterization.<locals>.draw_wrapper(artist, rendere
r, *args, **kwargs)
     36     renderer.stop_rasterizing()
     37     renderer._rasterizing = False
---> 39 return draw(artist, renderer, *args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/o
ffsetbox.py:383, in OffsetBox.draw(self, renderer)
    381 for c, (ox, oy) in zip(self.get_visible_children(), offsets):
    382     c.set_offset((px + ox, py + oy))
--> 383     c.draw(renderer)
    384 _bbox_artist(self, renderer, fill=False, props=dict(pad=0.))
    385 self.stale = False

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
rtist.py:39, in _prevent_rasterization.<locals>.draw_wrapper(artist, rendere
r, *args, **kwargs)
     36     renderer.stop_rasterizing()
     37     renderer._rasterizing = False
---> 39 return draw(artist, renderer, *args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/o
ffsetbox.py:383, in OffsetBox.draw(self, renderer)
    381 for c, (ox, oy) in zip(self.get_visible_children(), offsets):
    382     c.set_offset((px + ox, py + oy))
--> 383     c.draw(renderer)
    384 _bbox_artist(self, renderer, fill=False, props=dict(pad=0.))
    385 self.stale = False

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
rtist.py:39, in _prevent_rasterization.<locals>.draw_wrapper(artist, rendere
r, *args, **kwargs)
     36     renderer.stop_rasterizing()
     37     renderer._rasterizing = False
---> 39 return draw(artist, renderer, *args, **kwargs)
```

```
File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/offsetbox.py:379, in OffsetBox.draw(self, renderer)
    374 def draw(self, renderer):
    375     """
    376     Update the location of children if necessary and draw them
    377     to the given *renderer*.
    378     """
--> 379     bbox, offsets = self._get_bbox_and_child_offsets(renderer)
    380     px, py = self.get_offset(bbox, renderer)
    381     for c, (ox, oy) in zip(self.get_visible_children(), offsets):

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/offsetbox.py:452, in VPacker._get_bbox_and_child_offsets(self, renderer)
    449         if isinstance(c, PackerBase) and c.mode == "expand":
    450             c.set_width(self.width)
--> 452 bboxes = [c.get_bbox(renderer) for c in self.get_visible_children()]
    453 (x0, x1), xoffsets = _get_aligned_offsets(
    454     [bbox.intervalx for bbox in bboxes], self.width, self.align)
    455 height, yoffsets = _get_packed_offsets(
    456     [bbox.height for bbox in bboxes], self.height, sep, self.mode)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/offsetbox.py:452, in <listcomp>(.0)
    449         if isinstance(c, PackerBase) and c.mode == "expand":
    450             c.set_width(self.width)
--> 452 bboxes = [c.get_bbox(renderer) for c in self.get_visible_children()]
    453 (x0, x1), xoffsets = _get_aligned_offsets(
    454     [bbox.intervalx for bbox in bboxes], self.width, self.align)
    455 height, yoffsets = _get_packed_offsets(
    456     [bbox.height for bbox in bboxes], self.height, sep, self.mode)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/offsetbox.py:360, in OffsetBox.get_bbox(self, renderer)
    358 def get_bbox(self, renderer):
    359     """Return the bbox of the offsetbox, ignoring parent offsets."""
--> 360     bbox, offsets = self._get_bbox_and_child_offsets(renderer)
    361     return bbox

KeyboardInterrupt:
```

In [54]:
```
plot_returns(
    mean_returns, std_returns, method_name='dqn'
)
```

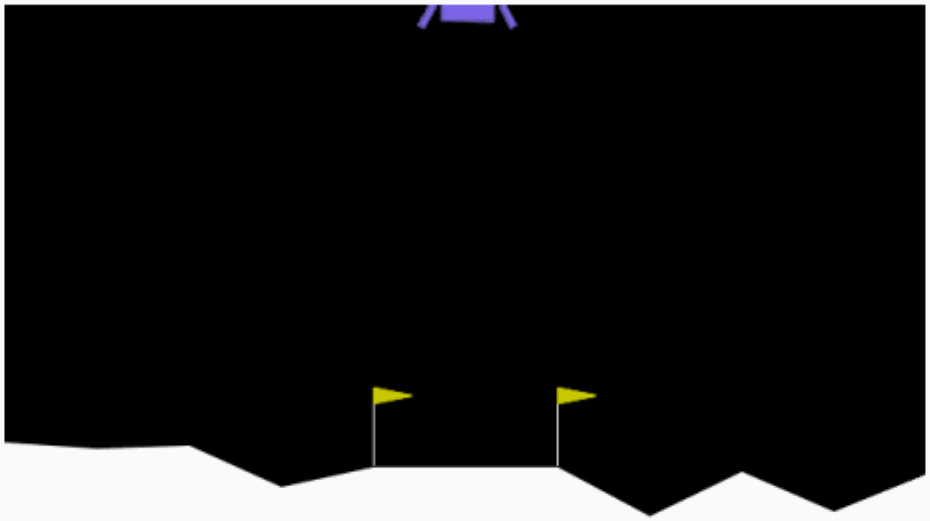



```
<Figure size 640x480 with 0 Axes>
<Figure size 640x480 with 0 Axes>
```

In [55]:
```
frames = demo_policy(
    Q, ENVIRONMENT_NAME
)
gif_path = save_frames_as_gif(frames, method_name='dqn')
Image(open(gif_path,'rb').read())
```

```
MovieWriter imagemagick unavailable; using Pillow instead.
```

Out[55]:



---

4. Actor-Critic

In [56]:
```
from src.networks import Policy
from src.networks import ValueFunctionQ
from src.ac import train_one_epoch as ac_epoch

################################### Hyper-parameters Tuning ################

EPOCHS: int = 1000
HIDDEN_DIMENSION: int = 64
LEARNING_RATE: float = 1e-3

############################################################################

# instantiate the state-action value function, Q
Q = ValueFunctionQ(
    state_dimension, num_actions, hidden_dimension=HIDDEN_DIMENSION
).to(DEVICE)
# initialize the target network
target_Q = ValueFunctionQ(
    state_dimension, num_actions, hidden_dimension=HIDDEN_DIMENSION
).to(DEVICE)
# initialize the policy network
policy = Policy(
    state_dimension, num_actions, hidden_dimension=HIDDEN_DIMENSION
).to(DEVICE)

# Learn the policy and Q
```

```
optimizer_Q = Adam(Q.parameters(), lr=LEARNING_RATE)
optimizer_pi = Adam(policy.parameters(), lr=LEARNING_RATE)

# memory.clear()
mean_returns, std_returns = [], []
for epoch in range(EPOCHS):

    # copy target network params
    target_Q.load_state_dict(Q.state_dict())

    ac_epoch(seed = SEED,
        env=environment,
        policy=policy, Q=Q, target_Q=target_Q,
        optimizer_Q=optimizer_Q, optimizer_pi=optimizer_pi
    )

    episode_return_mean, episode_return_std = eval_policy(policy, ENVIRONMEN
    mean_returns.append(episode_return_mean)
    std_returns.append(episode_return_std)

    logger.info(f'Epoch: {epoch:3d} \t return: {episode_return_mean:.2f}')

    if epoch:
        plot_returns(
            mean_returns, std_returns, method_name='ac-1-step-return', dynam
        )
```

```
In optimize_Q True False
2024-10-27 23:22:00.447 | INFO     | __main__:<module>:47 - Epoch: 236   ret
urn: -142.98
<Figure size 640x480 with 0 Axes>
```

```
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
Cell In[56], line 50
     47 logger.info(f'Epoch: {epoch:3d} \t return: {episode_return_mean:.2
f}')
     49 if epoch:
---> 50     plot_returns(
     51         mean_returns, std_returns, method_name='ac-1-step-return', d
ynamic=True
     52     )

File ~/Documents/Courses/CS_8803_[DRL]/cs8803drl-fall24/hw1/src/utils.py:12
4, in plot_returns(mean_returns, std_returns, method_name, dynamic)
    122 plt.title(method_name)
    123 plt.tight_layout()
--> 124 plt.show()
    126 plt.pause(0.01)  # pause a bit so that plots are updated
    127 display.display(plt.gcf())

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/p
yplot.py:612, in show(*args, **kwargs)
    568 """
    569 Display all open figures.
    570
   (...)
    609 explicitly there.
    610 """
    611 _warn_if_gui_out_of_main_thread()
--> 612 return _get_backend_mod().show(*args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib_i
nline/backend_inline.py:90, in show(close, block)
     88 try:
     89     for figure_manager in Gcf.get_all_fig_managers():
---> 90         display(
     91             figure_manager.canvas.figure,
     92             metadata=_fetch_figure_metadata(figure_manager.canvas.fi
gure)
     93         )
     94 finally:
     95     show._to_draw = []

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/display_functions.py:298, in display(include, exclude, metadata, transien
t, display_id, raw, clear, *objs, **kwargs)
    296     publish_display_data(data=obj, metadata=metadata, **kwargs)
    297 else:
--> 298     format_dict, md_dict = format(obj, include=include, exclude=excl
ude)
    299     if not format_dict:
    300         # nothing to display (e.g. _ipython_display_ took over)
    301         continue

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/formatters.py:182, in DisplayFormatter.format(self, obj, include, exclude)
    180 md = None
```

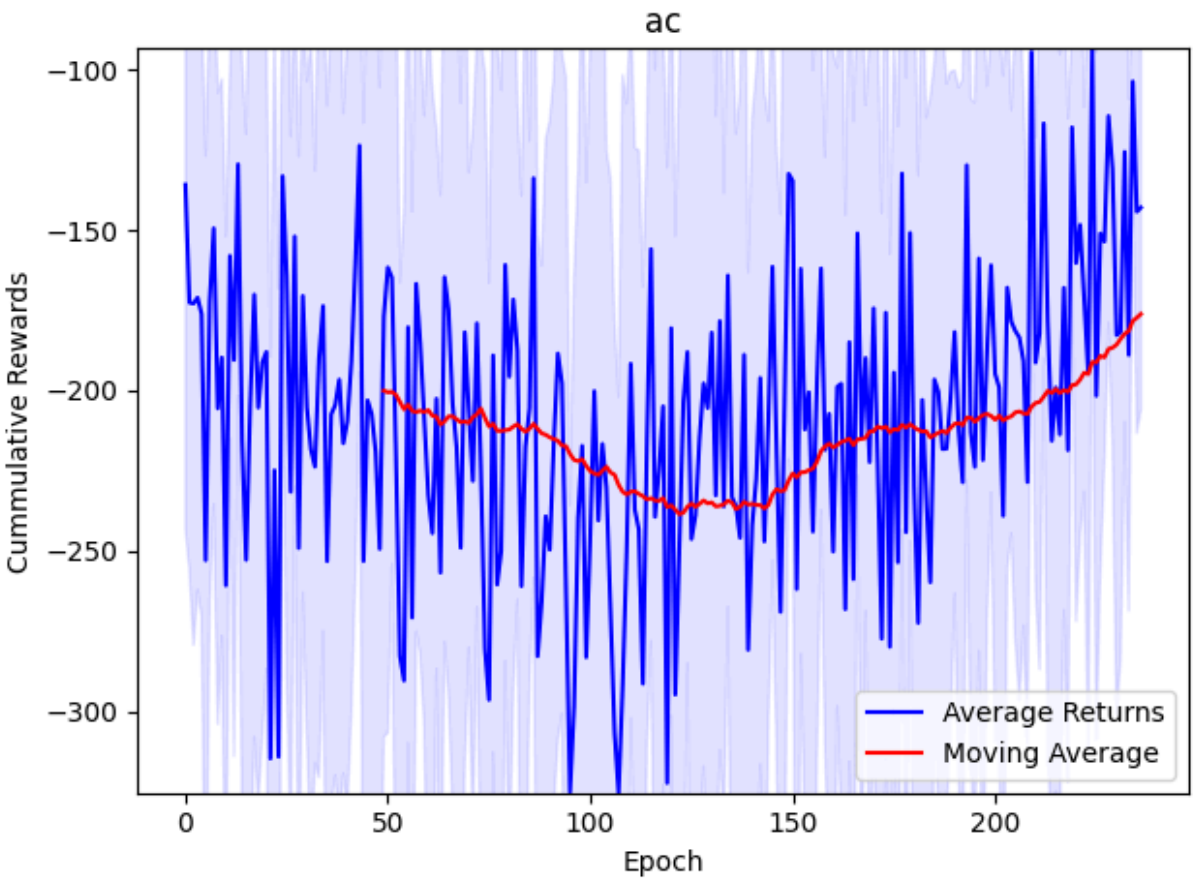
```
    181 try:
--> 182     data = formatter(obj)
    183 except:
    184     # FIXME: log the exception
    185     raise

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/decorator.p
y:232, in decorate.<locals>.fun(*args, **kw)
    230 if not kwsyntax:
    231     args, kw = fix(args, kw, sig)
--> 232 return caller(func, *(extras + args), **kw)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/formatters.py:226, in catch_format_error(method, self, *args, **kwargs)
    224 """show traceback on failed format call"""
    225 try:
--> 226     r = method(self, *args, **kwargs)
    227 except NotImplementedError:
    228     # don't warn on NotImplementedErrors
    229     return self._check_return(None, args[0])

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/formatters.py:343, in BaseFormatter.__call__(self, obj)
    341     pass
    342 else:
--> 343     return printer(obj)
    344 # Finally look for special method names
    345 method = get_real_method(obj, self.print_method)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/IPython/cor
e/pylabtools.py:170, in print_figure(fig, fmt, bbox_inches, base64, **kwarg
s)
    167     from matplotlib.backend_bases import FigureCanvasBase
    168     FigureCanvasBase(fig)
--> 170 fig.canvas.print_figure(bytes_io, **kw)
    171 data = bytes_io.getvalue()
    172 if fmt == 'svg':

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/b
ackend_bases.py:2178, in FigureCanvasBase.print_figure(self, filename, dpi,
facecolor, edgecolor, orientation, format, bbox_inches, pad_inches, bbox_ext
ra_artists, backend, **kwargs)
   2176 if bbox_inches:
   2177     if bbox_inches == "tight":
-> 2178         bbox_inches = self.figure.get_tightbbox(
   2179             renderer, bbox_extra_artists=bbox_extra_artists)
   2180         if (isinstance(layout_engine, ConstrainedLayoutEngine) and
   2181                 pad_inches == "layout"):
   2182             h_pad = layout_engine.get()["h_pad"]

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/_
api/deprecation.py:457, in make_keyword_only.<locals>.wrapper(*args, **kwarg
s)
    451 if len(args) > name_idx:
    452     warn_deprecated(
    453         since, message="Passing the %(name)s %(obj_type)s "
```

```
    454         "positionally is deprecated since Matplotlib %(since)s; the
"
    455         "parameter will become keyword-only %(removal)s.",
    456         name=name, obj_type=f"parameter of {func.__name__}()")
--> 457 return func(*args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/f
igure.py:1787, in FigureBase.get_tightbbox(self, renderer, bbox_extra_artist
s)
   1783 if ax.get_visible():
   1784     # some Axes don't take the bbox_extra_artists kwarg so we
   1785     # need this conditional....
   1786     try:
-> 1787         bbox = ax.get_tightbbox(
   1788             renderer, bbox_extra_artists=bbox_extra_artists)
   1789     except TypeError:
   1790         bbox = ax.get_tightbbox(renderer)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/_
api/deprecation.py:457, in make_keyword_only.<locals>.wrapper(*args, **kwarg
s)
    451 if len(args) > name_idx:
    452     warn_deprecated(
    453         since, message="Passing the %(name)s %(obj_type)s "
    454         "positionally is deprecated since Matplotlib %(since)s; the
"
    455         "parameter will become keyword-only %(removal)s.",
    456         name=name, obj_type=f"parameter of {func.__name__}()")
--> 457 return func(*args, **kwargs)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
xes/_base.py:4479, in _AxesBase.get_tightbbox(self, renderer, call_axes_loca
tor, bbox_extra_artists, for_layout_only)
   4477         if ba:
   4478             bb.append(ba)
-> 4479 self._update_title_position(renderer)
   4480 axbbox = self.get_window_extent(renderer)
   4481 bb.append(axbbox)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
xes/_base.py:3045, in _AxesBase._update_title_position(self, renderer)
   3043 top = max(top, bb.ymax)
   3044 if title.get_text():
-> 3045     ax.yaxis.get_tightbbox(renderer)  # update offsetText
   3046     if ax.yaxis.offsetText.get_text():
   3047         bb = ax.yaxis.offsetText.get_tightbbox(renderer)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
xis.py:1372, in Axis.get_tightbbox(self, renderer, for_layout_only)
   1369     renderer = self.figure._get_renderer()
   1370 ticks_to_draw = self._update_ticks()
-> 1372 self._update_label_position(renderer)
   1374 # go back to just this axis's tick labels
   1375 tlb1, tlb2 = self._get_ticklabel_bboxes(ticks_to_draw, renderer)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
```

```
xis.py:2654, in YAxis._update_label_position(self, renderer)
   2650     return
   2652 # get bounding boxes for this axis and any siblings
   2653 # that have been set by `fig.align_ylabels()`
-> 2654 bboxes, bboxes2 = self._get_tick_boxes_siblings(renderer=renderer)
   2655 x, y = self.label.get_position()
   2656 if self.label_position == 'left':

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
xis.py:2206, in Axis._get_tick_boxes_siblings(self, renderer)
   2204 axis = ax._axis_map[name]
   2205 ticks_to_draw = axis._update_ticks()
-> 2206 tlb, tlb2 = axis._get_ticklabel_bboxes(ticks_to_draw, renderer)
   2207 bboxes.extend(tlb)
   2208 bboxes2.extend(tlb2)

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
xis.py:1351, in Axis._get_ticklabel_bboxes(self, ticks, renderer)
   1349 if renderer is None:
   1350     renderer = self.figure._get_renderer()
-> 1351 return ([tick.label1.get_window_extent(renderer)
   1352          for tick in ticks if tick.label1.get_visible()],
   1353         [tick.label2.get_window_extent(renderer)
   1354          for tick in ticks if tick.label2.get_visible()])

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/a
xis.py:1351, in <listcomp>(.0)
   1349 if renderer is None:
   1350     renderer = self.figure._get_renderer()
-> 1351 return ([tick.label1.get_window_extent(renderer)
   1352          for tick in ticks if tick.label1.get_visible()],
   1353         [tick.label2.get_window_extent(renderer)
   1354          for tick in ticks if tick.label2.get_visible()])

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/t
ext.py:959, in Text.get_window_extent(self, renderer, dpi)
    954     raise RuntimeError(
    955         "Cannot get window extent of text w/o renderer. You likely "
    956         "want to call 'figure.draw_without_rendering()' first.")
    958 with cbook._setattr_cm(self.figure, dpi=dpi):
--> 959     bbox, info, descent = self._get_layout(self._renderer)
    960     x, y = self.get_unitless_position()
    961     x, y = self.get_transform().transform((x, y))

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/t
ext.py:422, in Text._get_layout(self, renderer)
    419 ymin = ys[-1] - descent  # baseline of last line minus its descent
    421 # get the rotation matrix
--> 422 M = Affine2D().rotate_deg(self.get_rotation())
    424 # now offset the individual text lines within the box
    425 malign = self._get_multialignment()

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/t
ransforms.py:1909, in Affine2D.__init__(self, matrix, **kwargs)
   1899 def __init__(self, matrix=None, **kwargs):
   1900     """
```

```
   1901     Initialize an Affine transform from a 3x3 numpy float array::
   1902
   (...)
   1907     If *matrix* is None, initialize with the identity transform.
   1908     """
-> 1909     super().__init__(**kwargs)
   1910     if matrix is None:
   1911         # A bit faster than np.identity(3).
   1912         matrix = IdentityTransform._mtx

File ~/anaconda3/envs/cs8803drl_A1/lib/python3.10/site-packages/matplotlib/transforms.py:1780, in AffineBase.__init__(self, *args, **kwargs)
   1779 def __init__(self, *args, **kwargs):
-> 1780     super().__init__(*args, **kwargs)
   1781     self._inverted = None

KeyboardInterrupt:
```

In [57]:
```
plot_returns(
    mean_returns, std_returns, method_name='ac'
)
```




```
<Figure size 640x480 with 0 Axes>
<Figure size 640x480 with 0 Axes>
```

In [58]:
```
frames = demo_policy(
    policy, ENVIRONMENT_NAME,
)
```

```
gif_path = save_frames_as_gif(frames, method_name='ac')
Image(open(gif_path,'rb').read())
```

```
MovieWriter imagemagick unavailable; using Pillow instead.
```

Out[58]: