Continuum Robot PPO report

| Model Name	./model/reward_f2
Time	2025-04-20 20:36
Model_Type	PPO
Timesteps	22000000
Control Mode	1
Device	cuda
Network Arch	[1024, 1024, 512]
Average Error	0.6604152043157897
batch	16384
buffer_size	100
train_freq	1
learning_starts	1
n_steps	2048
n_epochs	10
learning_rate	0.0003
n_env	48
best_value_loss	0.4016168554623922

Reward Function Source

```
def my_custom_reward1(distance):
    return -distance + np.exp(-distance**2) * 5
```
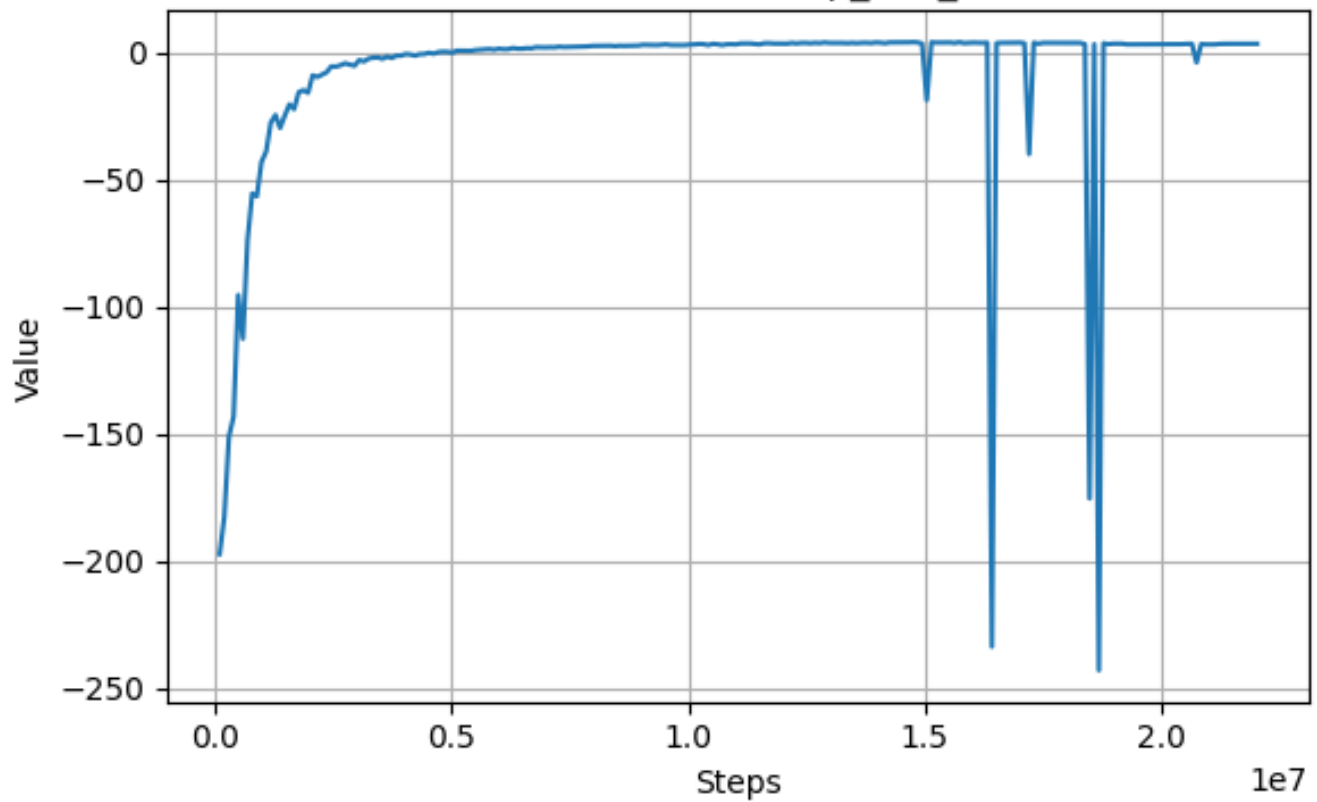
Done Function Source

```
def my_custom_done1(reward, step):
    return reward>-2
```
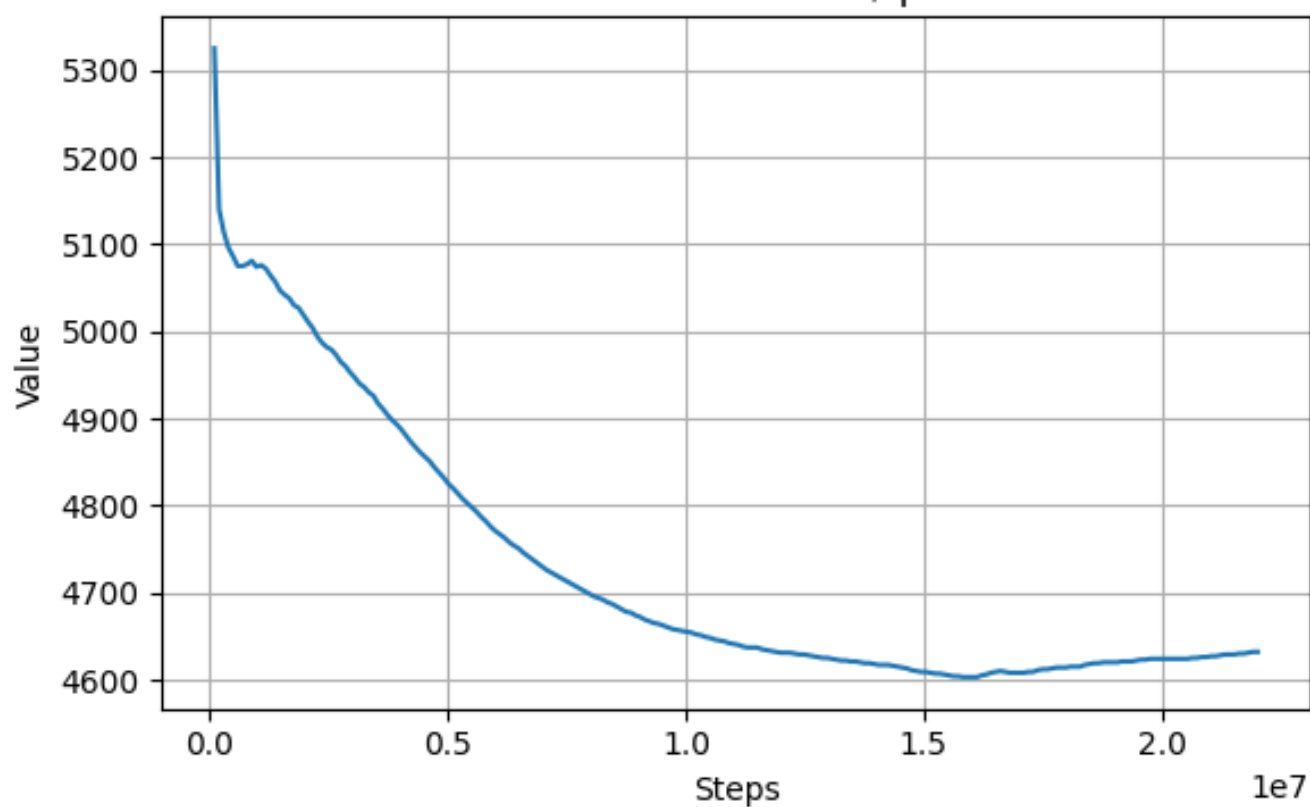
Training Reward Curve
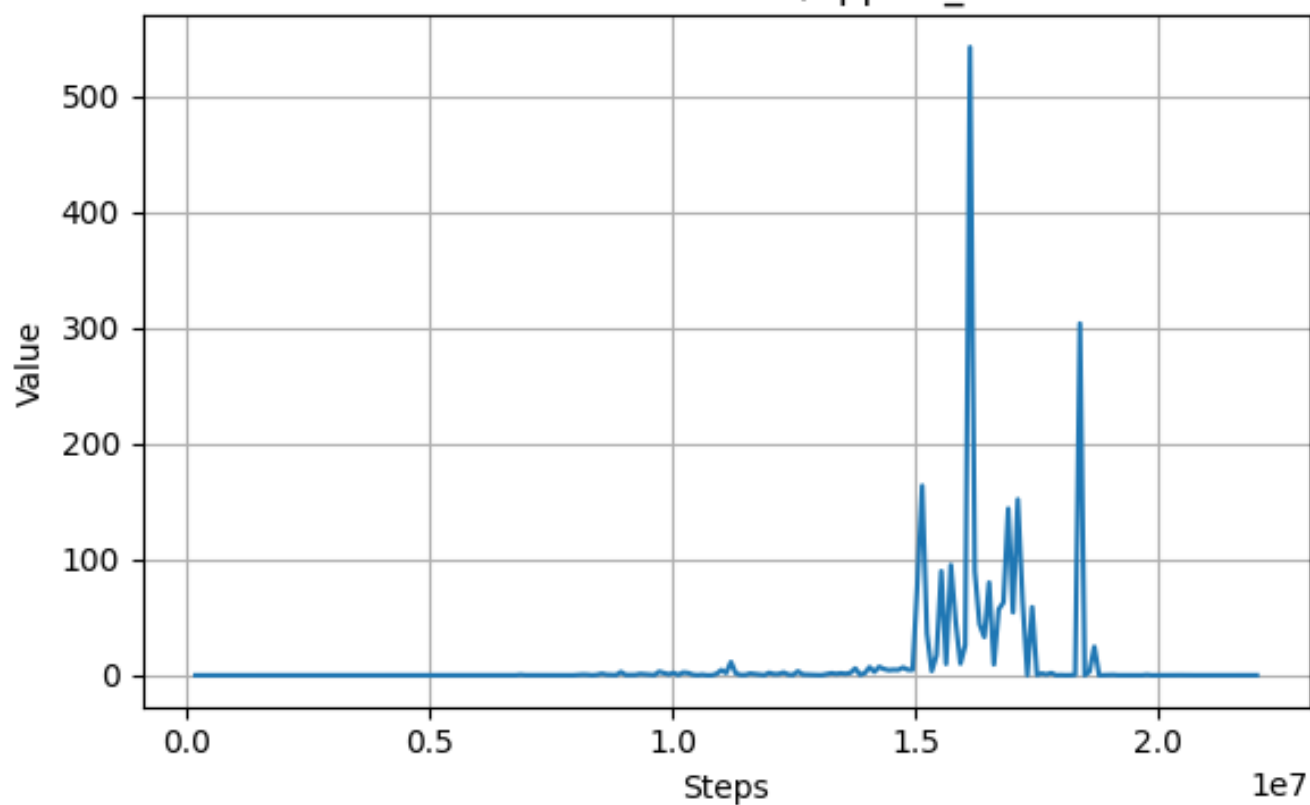
TensorBoard: rollout/ep_len_mean
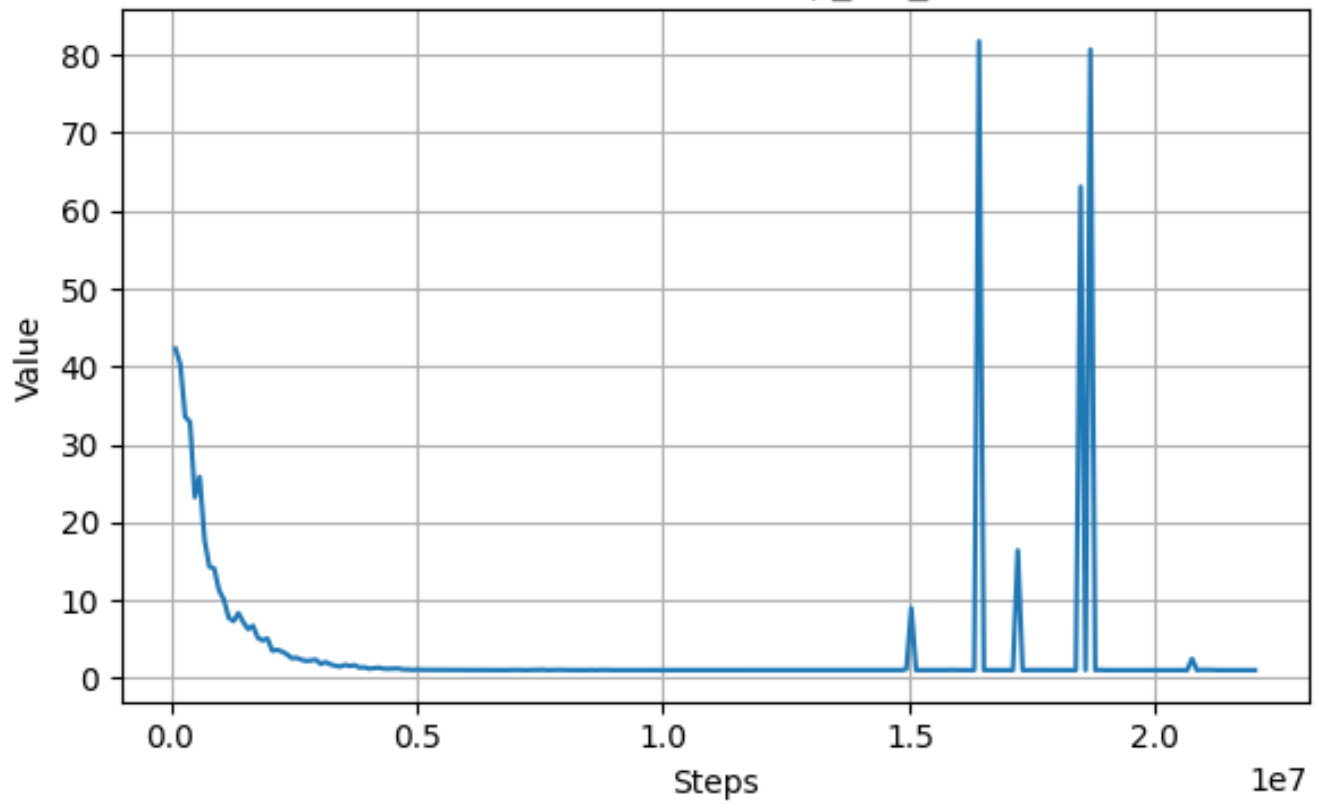
TensorBoard: rollout/ep_rew_mean
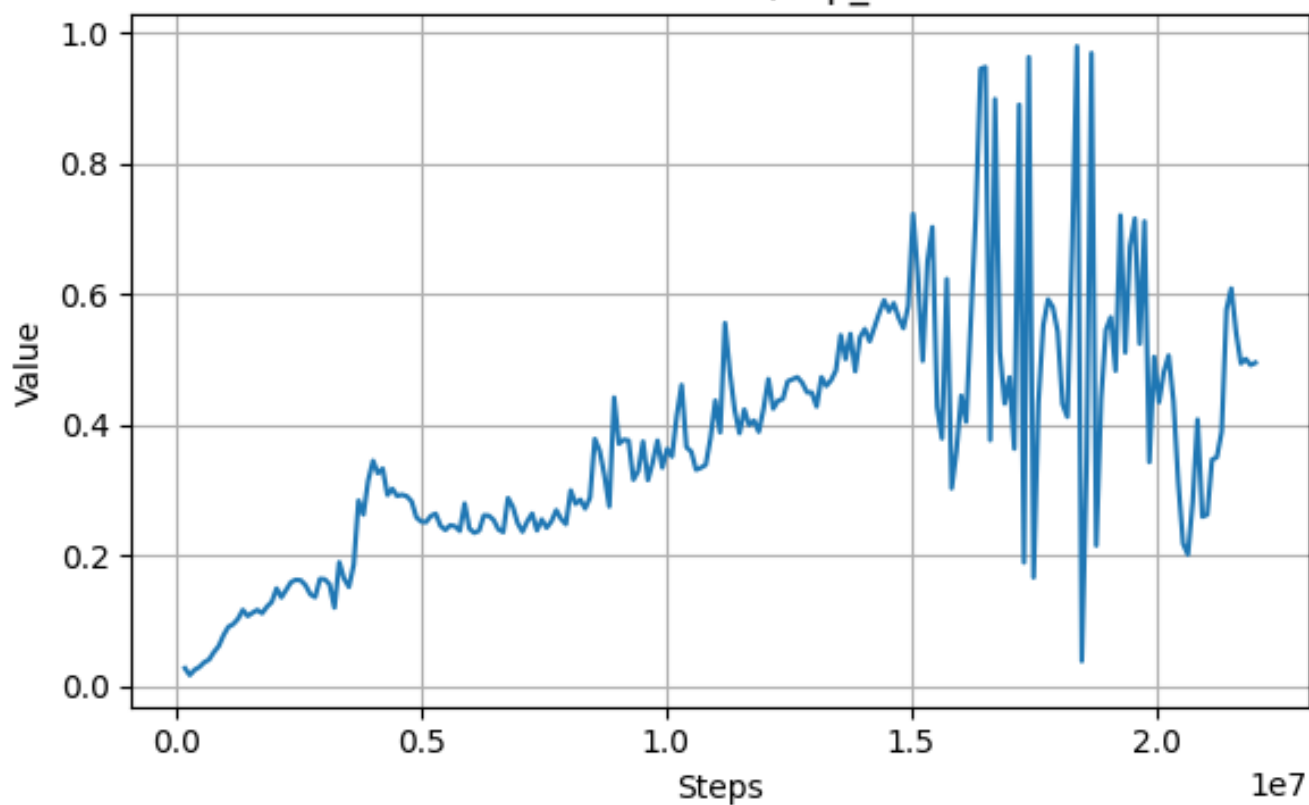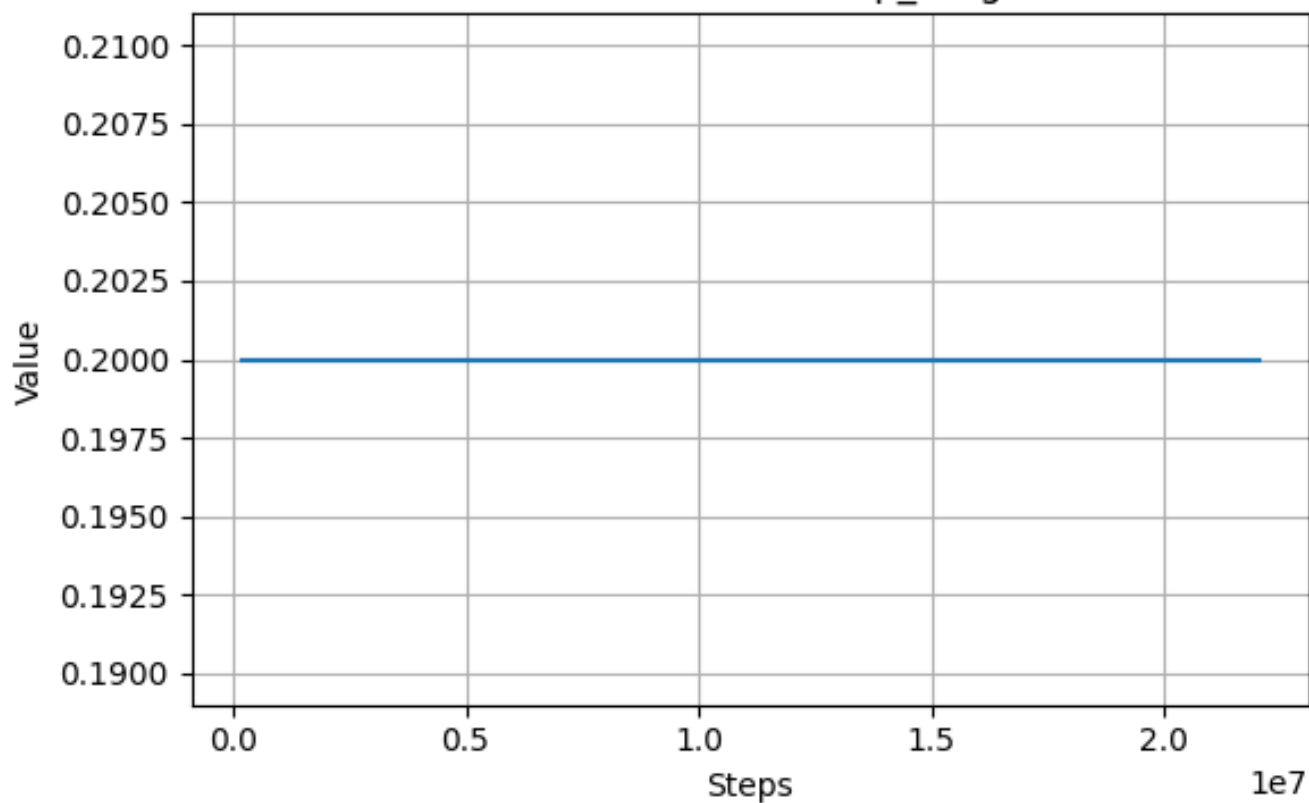
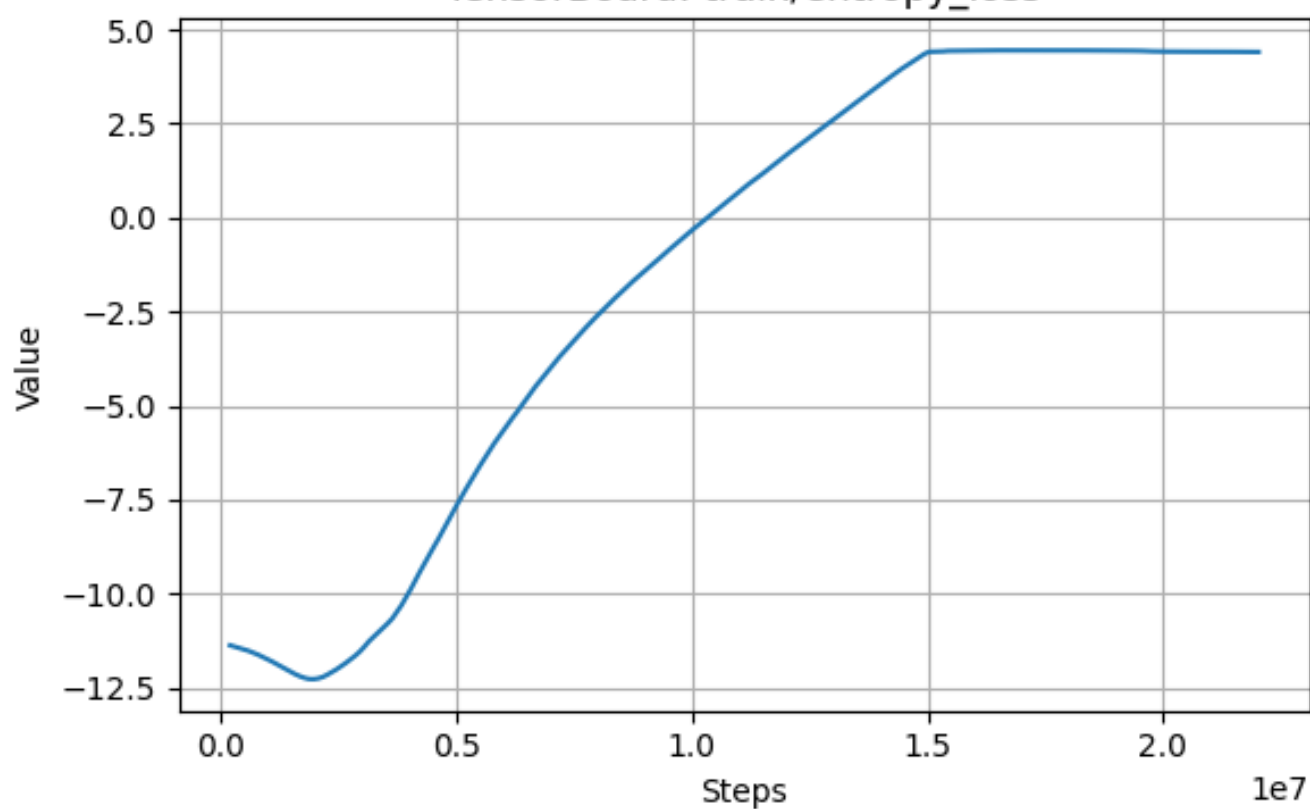TensorBoard: time/fps

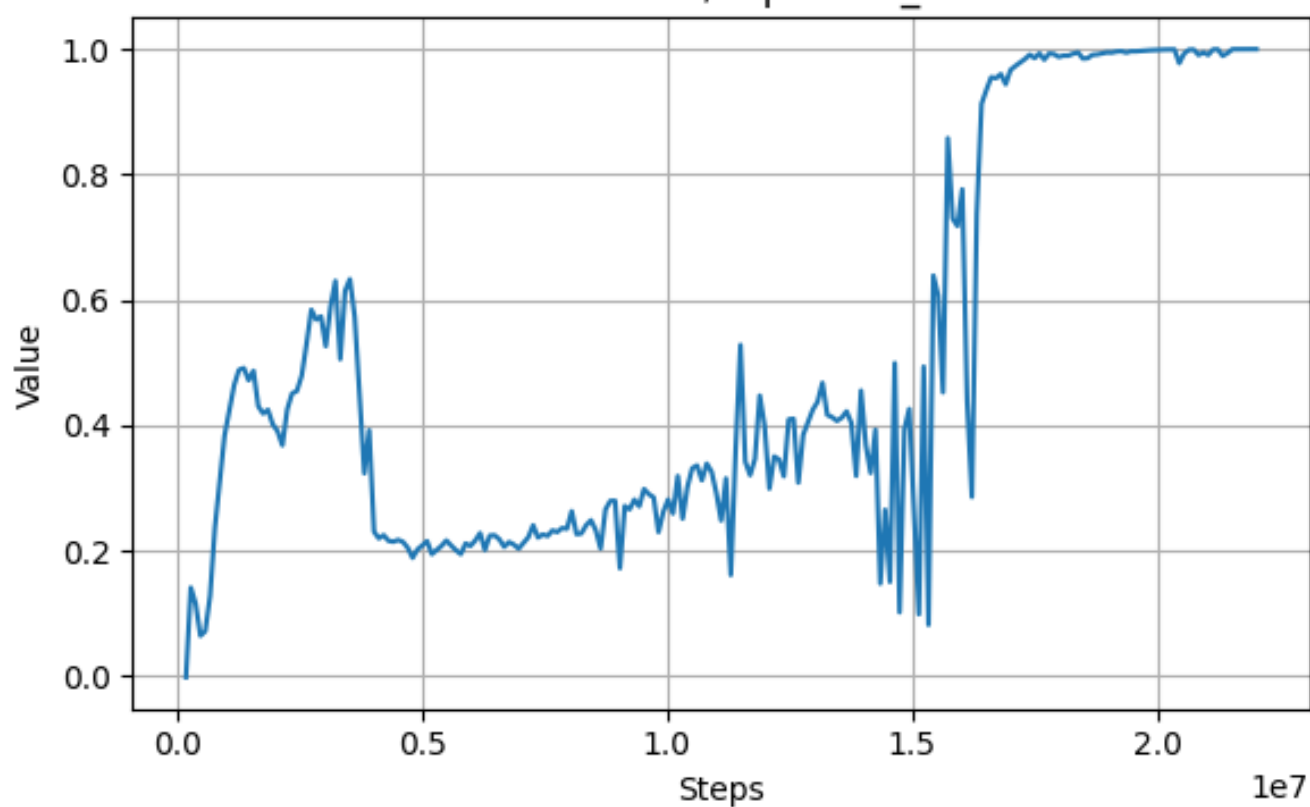TensorBoard: train/approx_kl

TensorBoard: train/clip_fraction
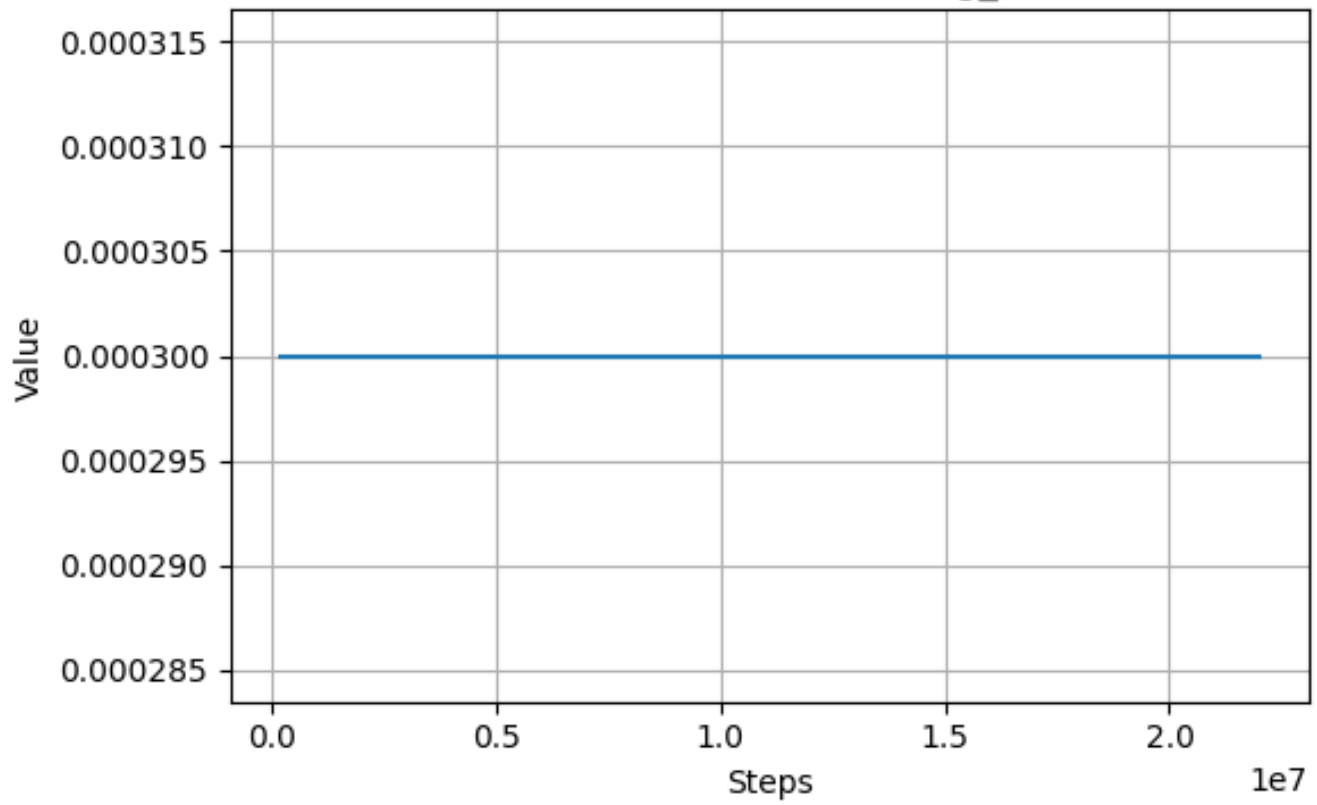
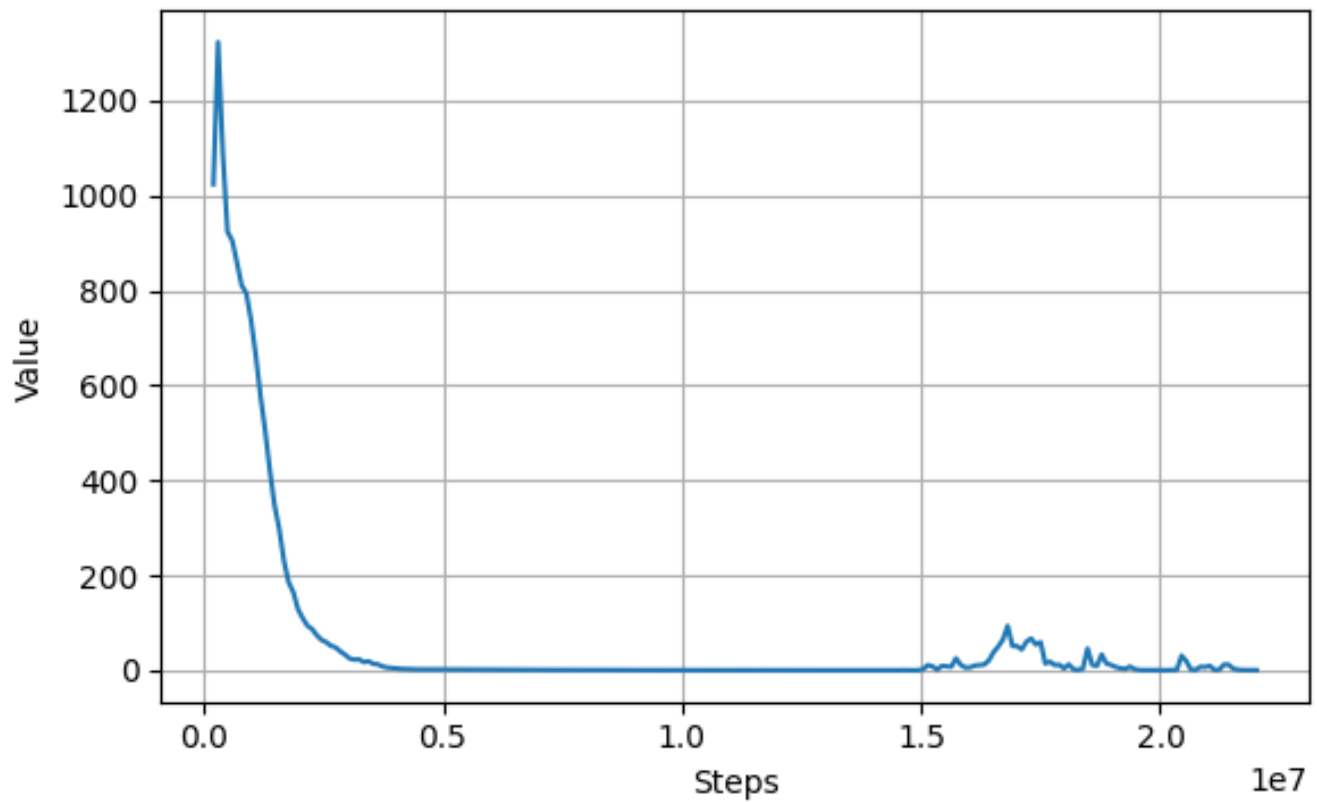TensorBoard: train/clip_range

TensorBoard: train/entropy_loss

TensorBoard: train/explained_variance

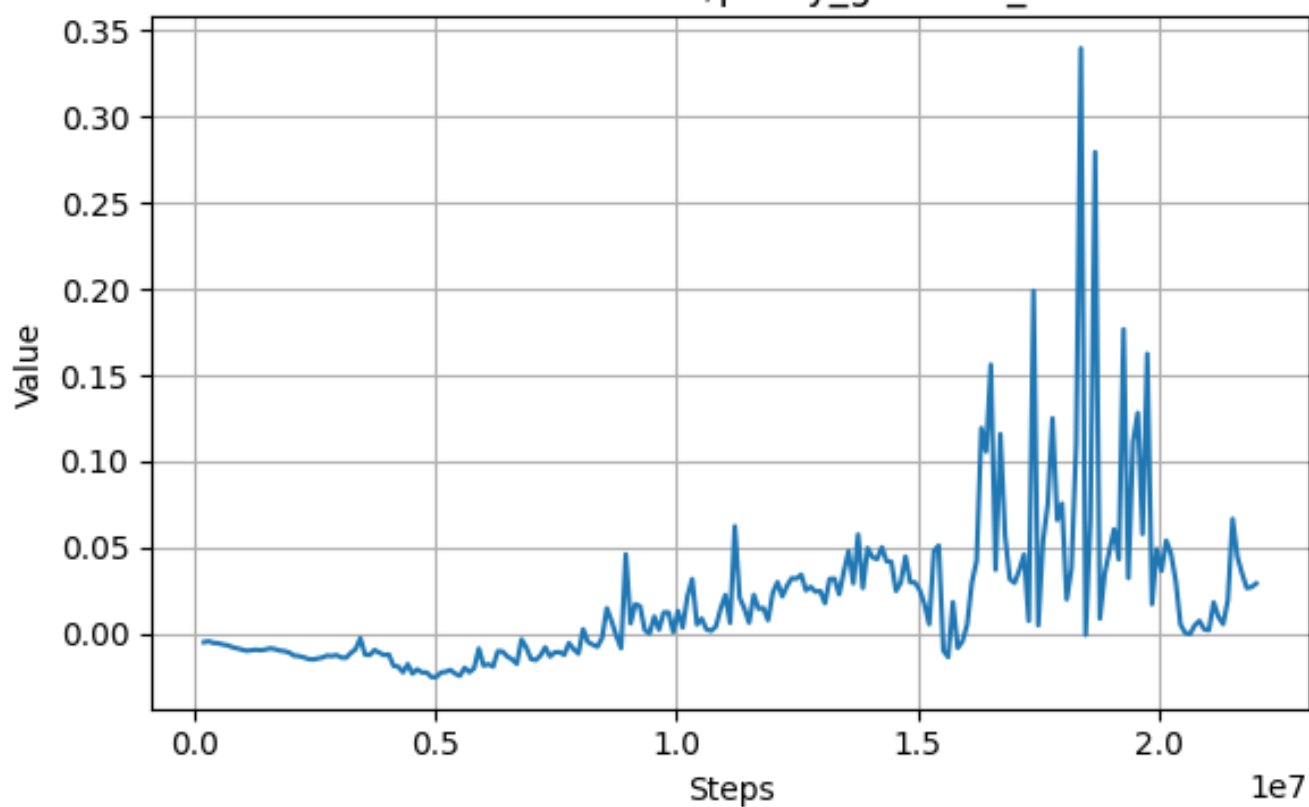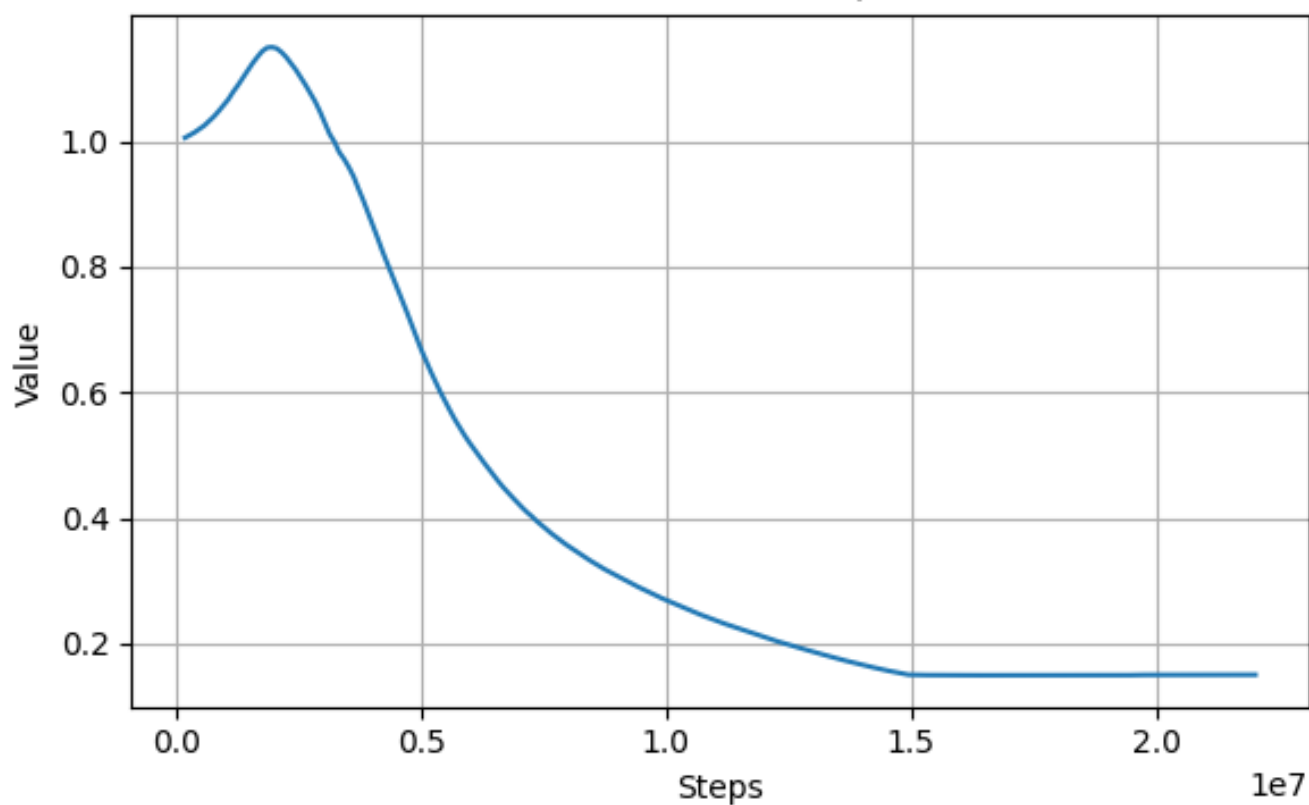TensorBoard: train/learning_rate

TensorBoard: train/loss

TensorBoard: train/policy_gradient_loss

TensorBoard: train/std

TensorBoard: train/value_loss