

Continuum Robot PPO report

Model Name	./model/test_1
Time	2025-04-22 19:17
Model_Type	PPO
seed	1
Timesteps	1000000
Control Mode	1
Device	cuda
Network Arch	[1024, 1024, 512]
Average Error	3.509127838809671
batch	16384
buffer_size	100000
train_freq	4
learning_starts	20000
n_steps	4096
n_epochs	10
learning_rate	0.0003
n_env	24
best_reward	-0.36961742876237197

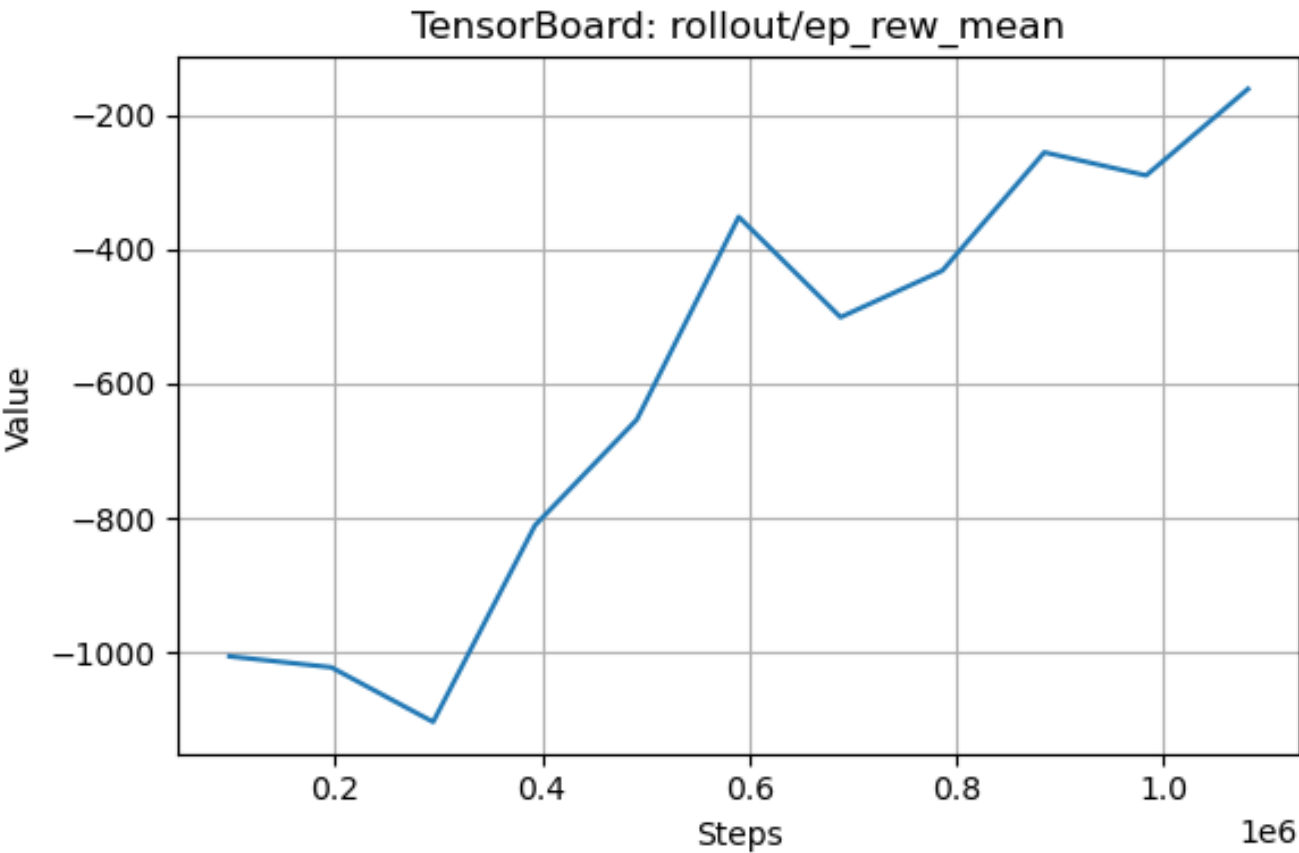
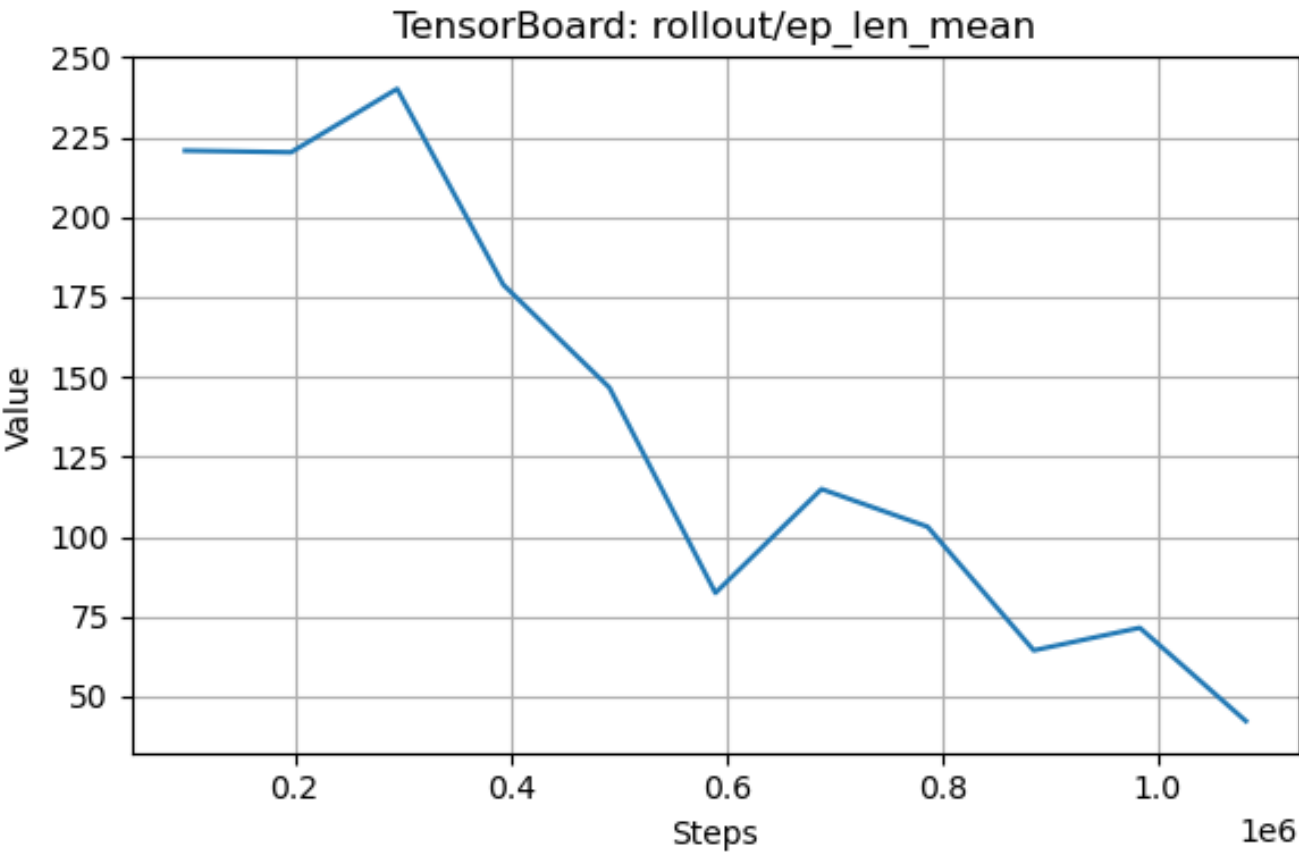
Reward Function Source

```
def my_custom_reward(distance):  
    return -distance + np.exp(-distance**2) * 5
```

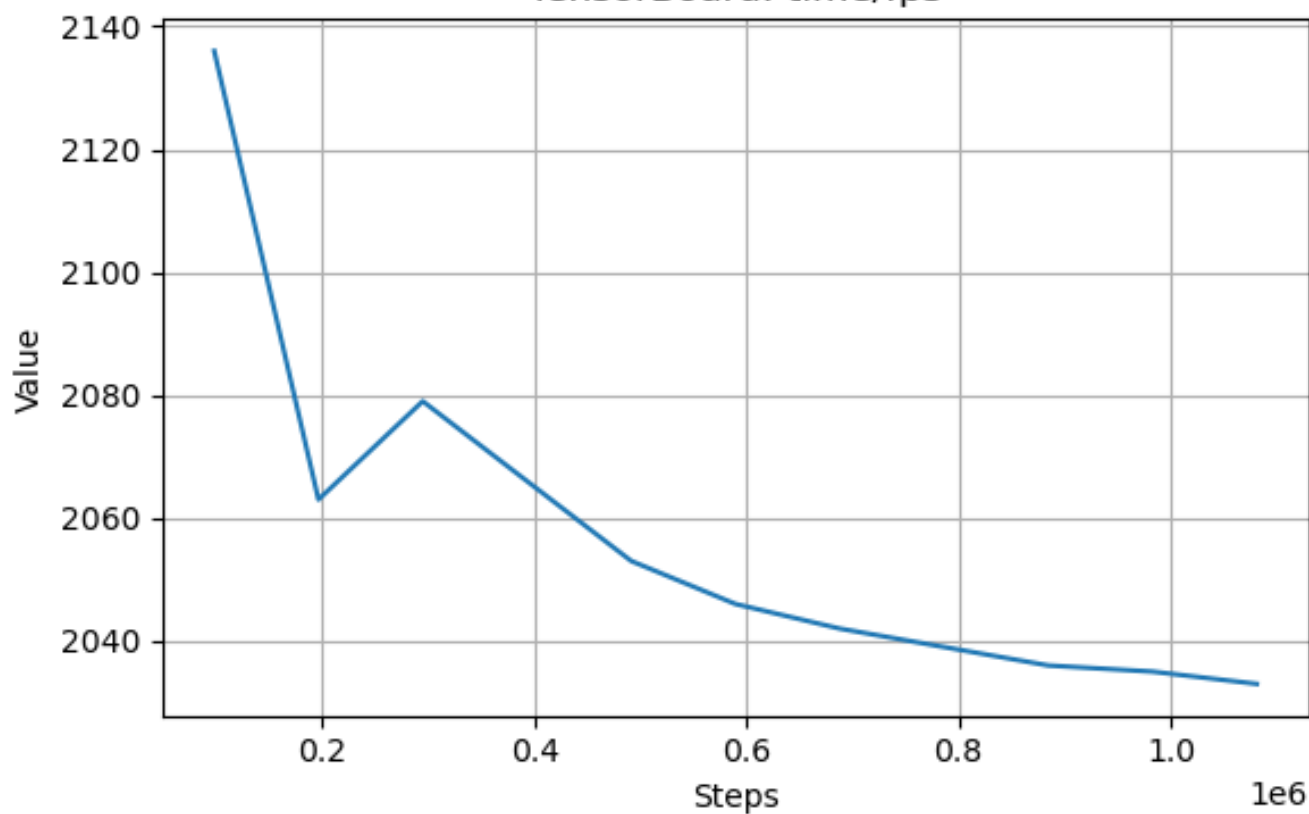
Done Function Source

```
def my_custom_done(reward, step,distance,in_step):  
    return reward>=-0.5
```

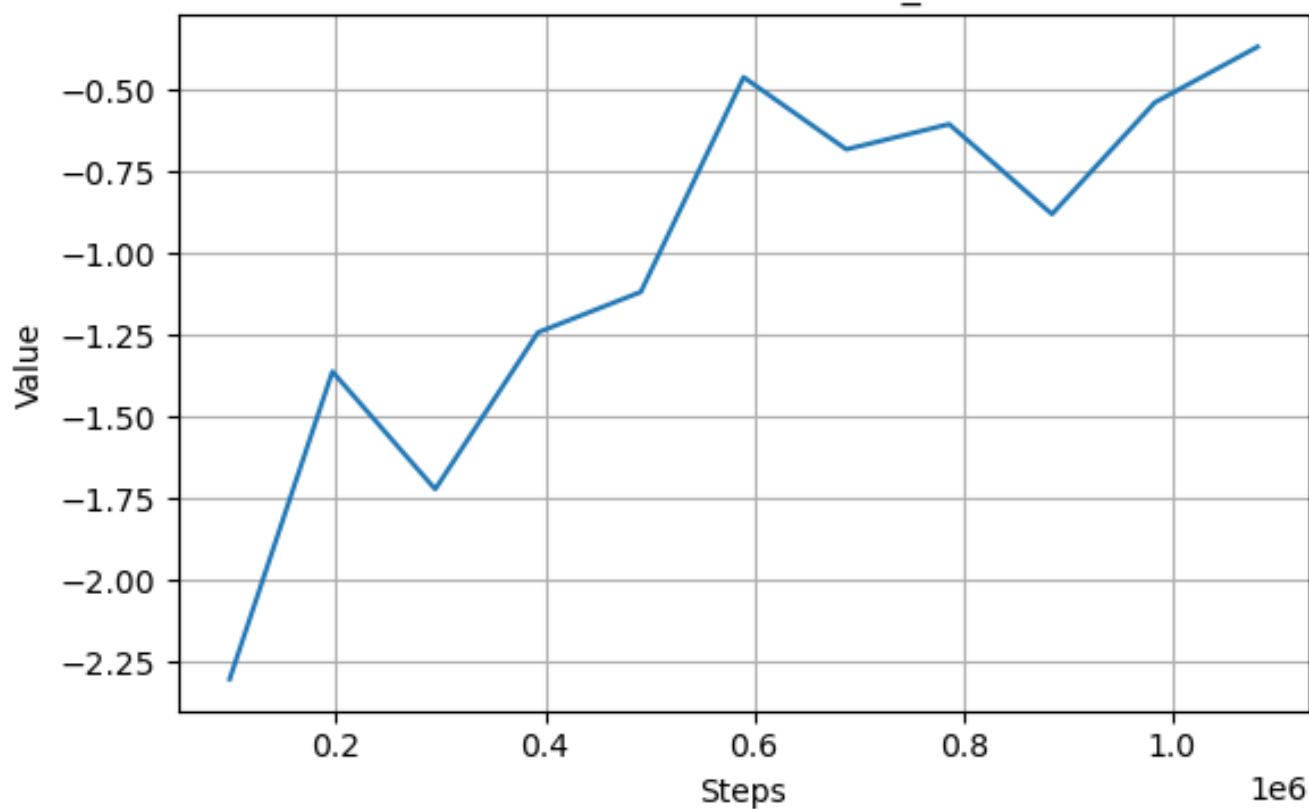
Training Reward Curve



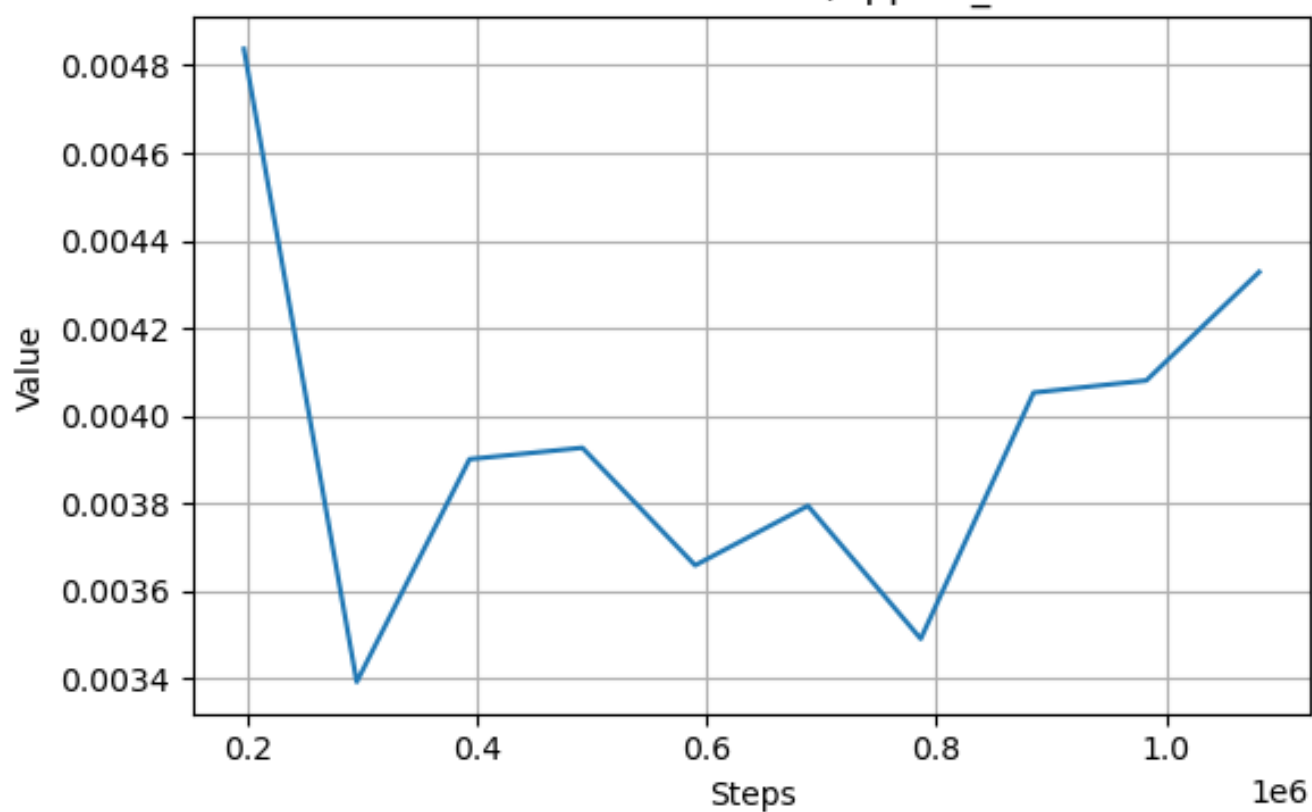
TensorBoard: time/fps



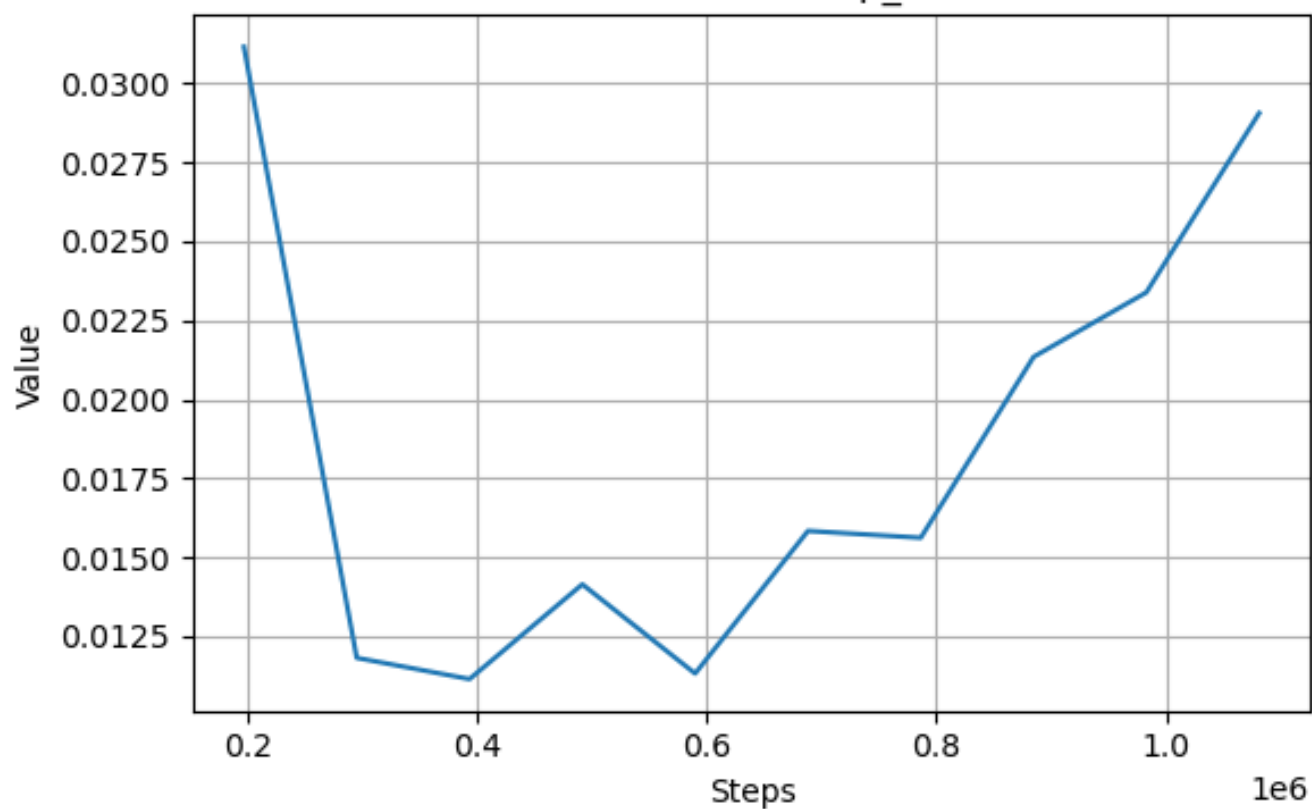
TensorBoard: train/mean_reward



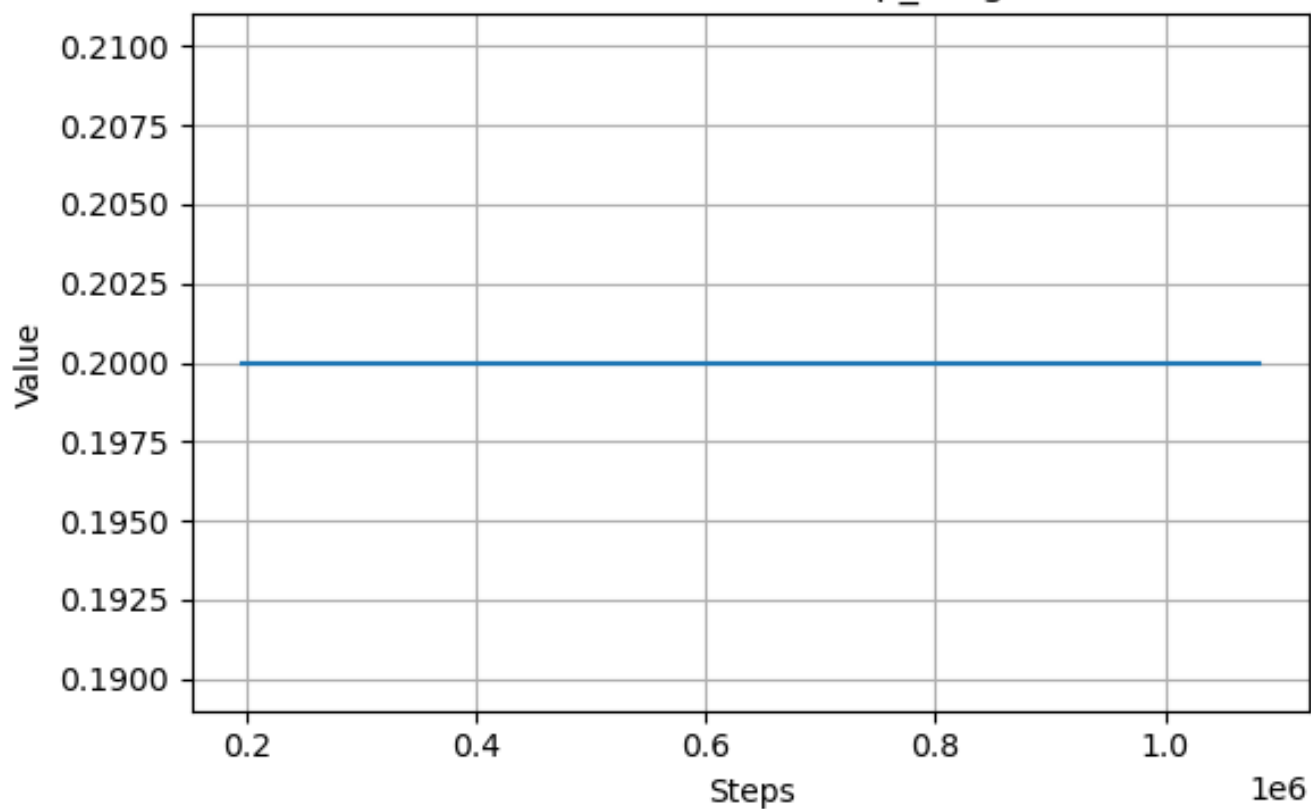
TensorBoard: train/approx_kl



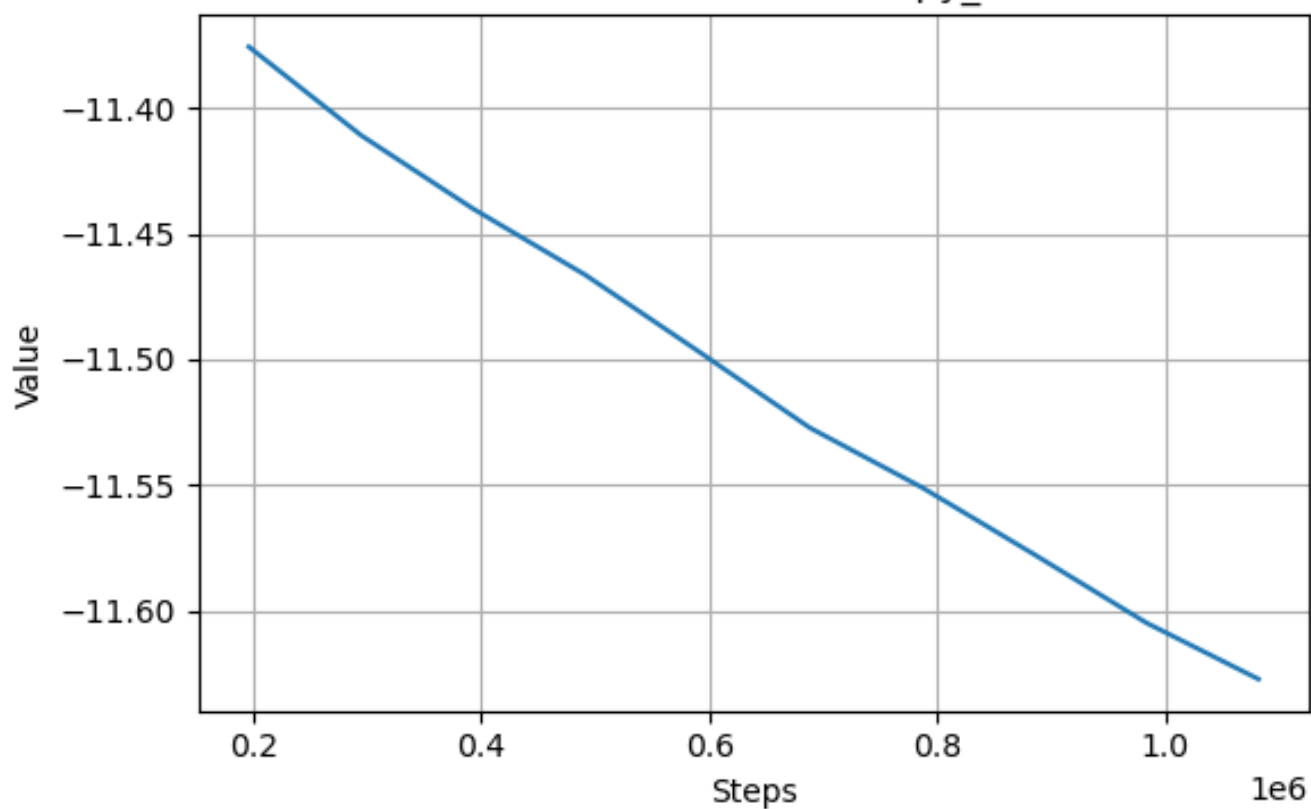
TensorBoard: train/clip_fraction



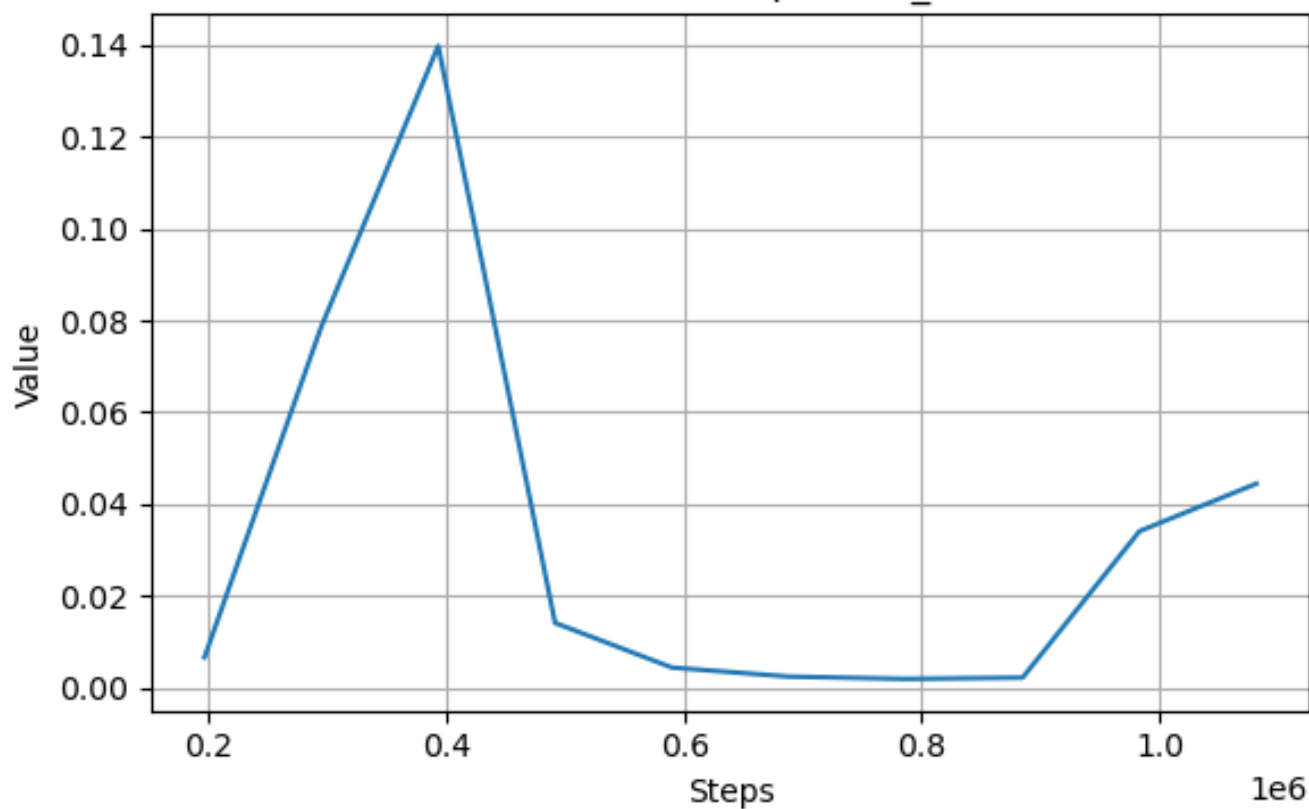
TensorBoard: train/clip_range



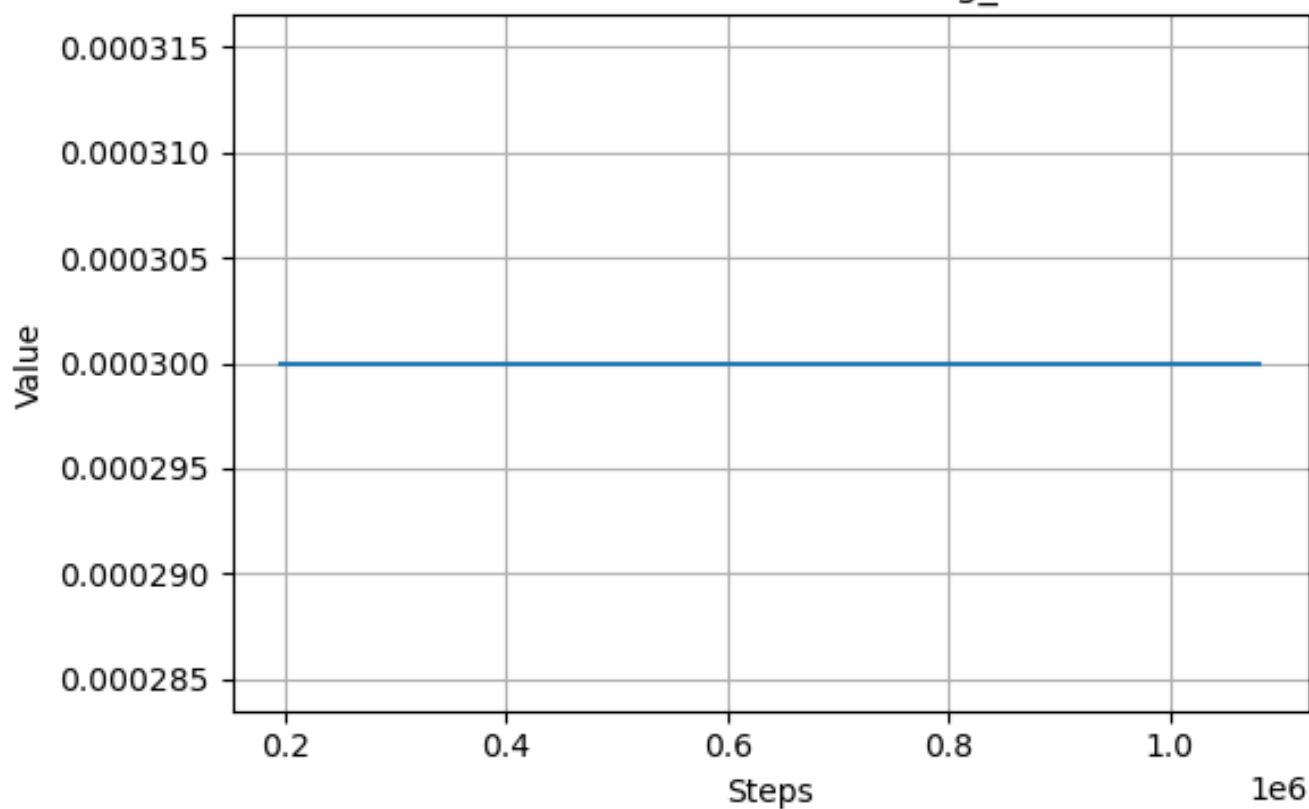
TensorBoard: train/entropy_loss



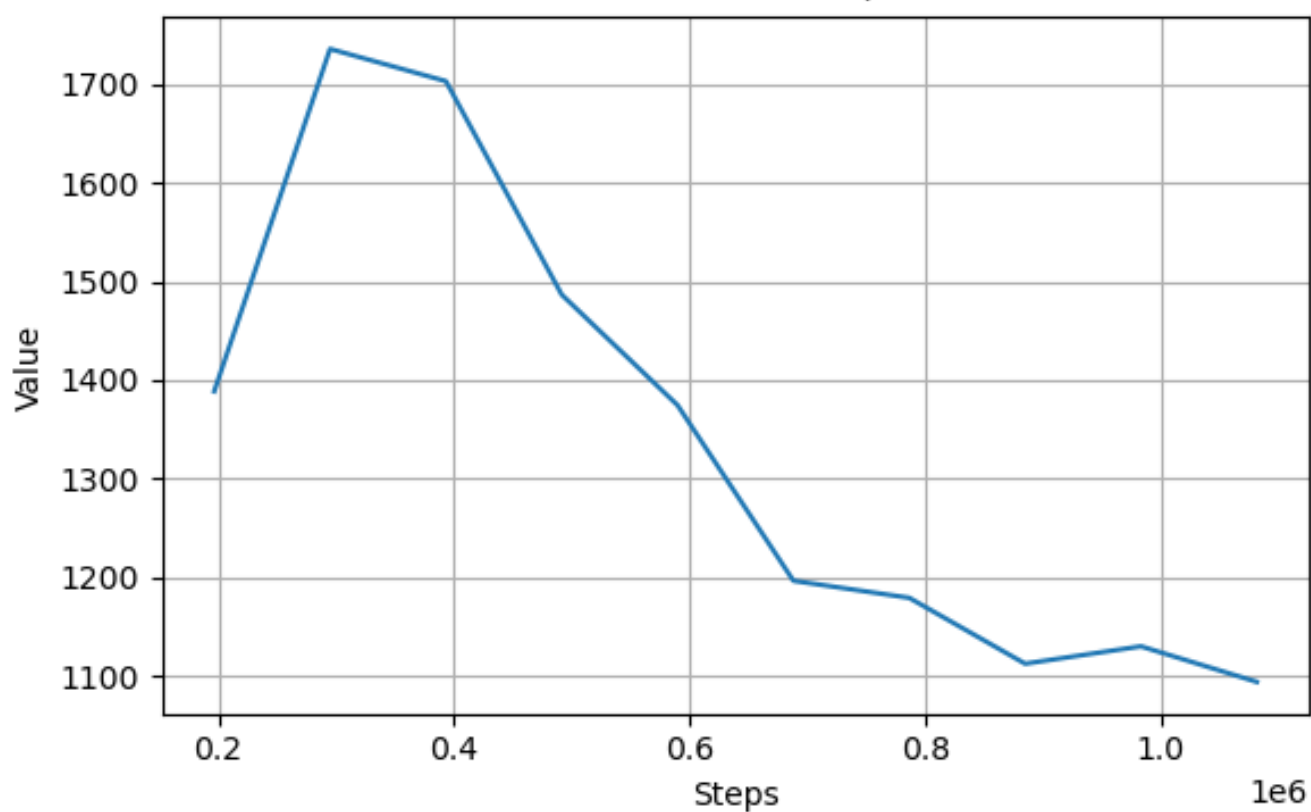
TensorBoard: train/explained_variance



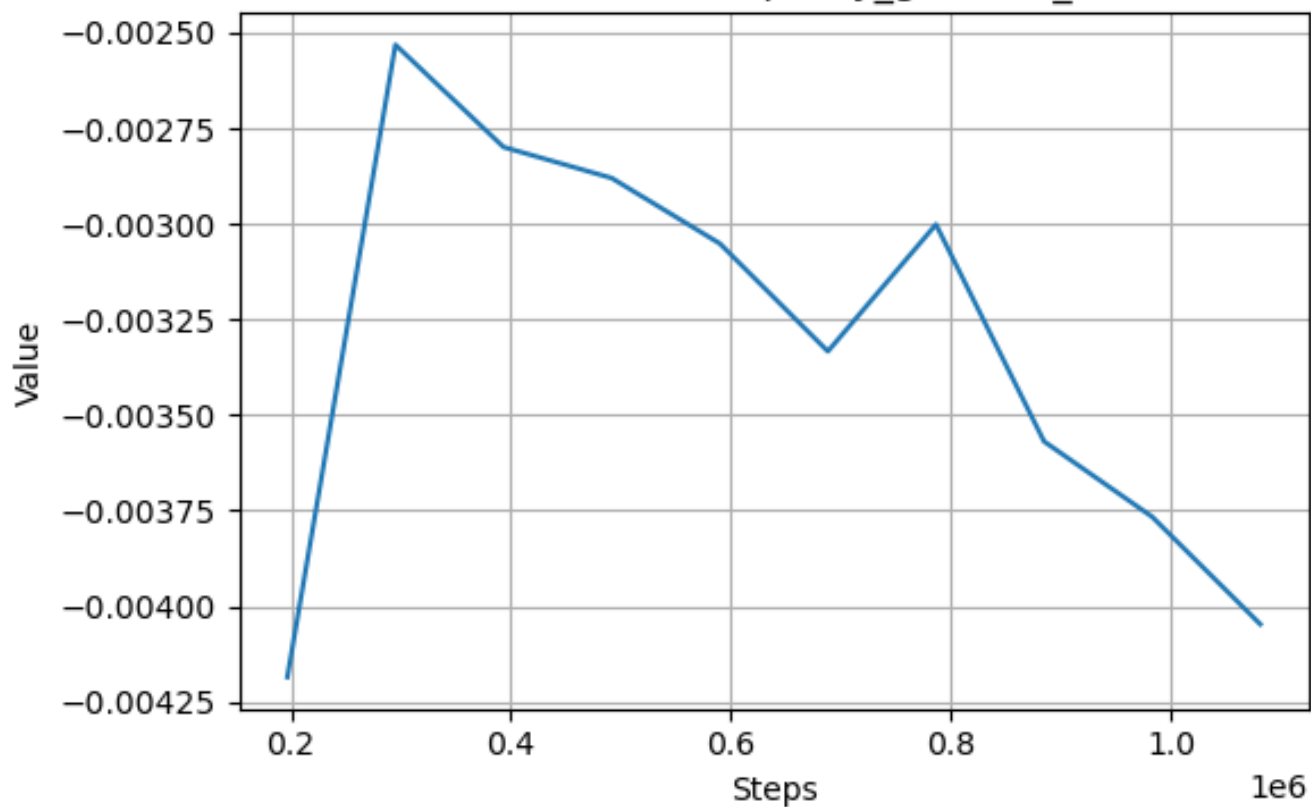
TensorBoard: train/learning_rate



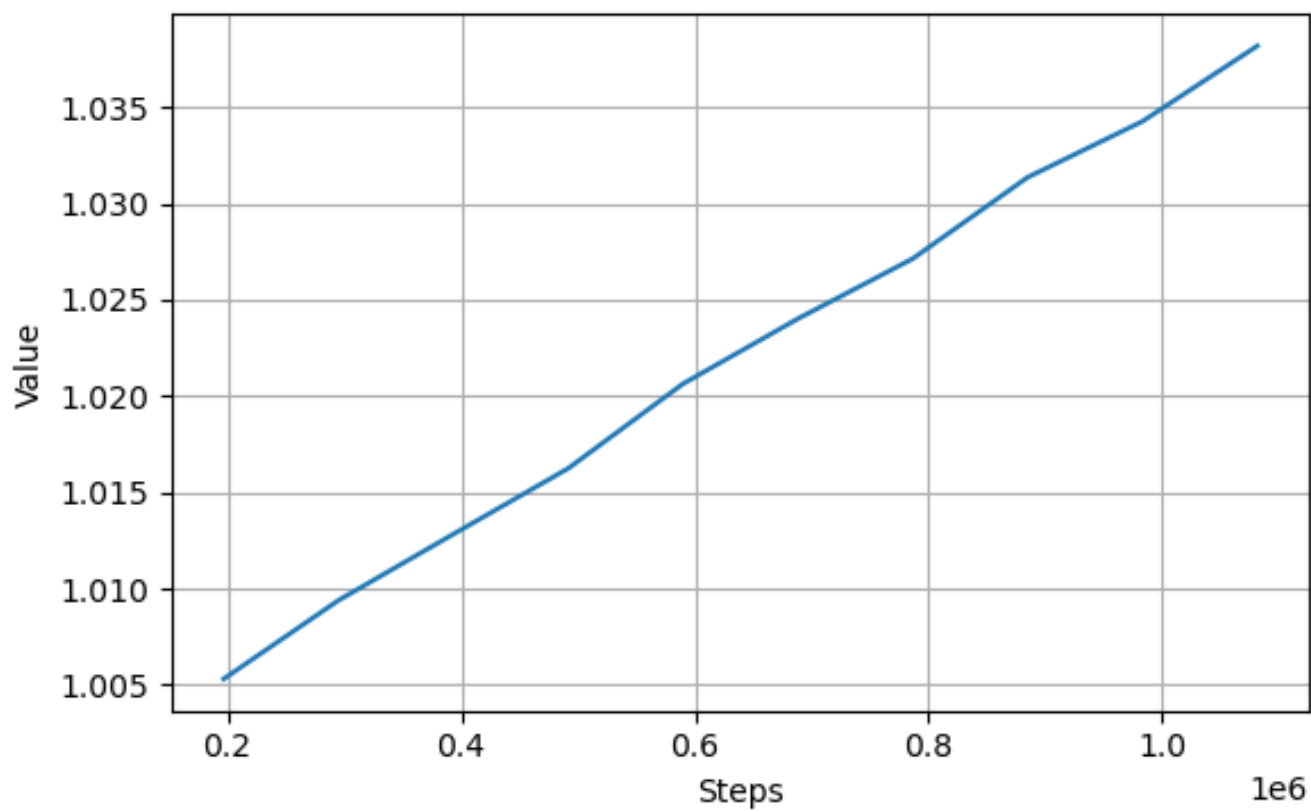
TensorBoard: train/loss



TensorBoard: train/policy_gradient_loss



TensorBoard: train/std



TensorBoard: train/value_loss

