

PPO: Performance with tuned hyperparameters and cosine learning rate annealing

