第三次作业

1. 命令行执行代码

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name PongNoFrameskip-v4 --exp\_name q1

生成run-log文件后可视化遇到问题，未解决

1. 对比DQN和DoubleDQN，命令行执行如下代码：

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name LunarLander-v2 --exp\_name q2\_dqn\_1 --seed 1

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name LunarLander-v2 --exp\_name q2\_dqn\_2 --seed 2

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name LunarLander-v2 --exp\_name q2\_dqn\_3 --seed 3

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name LunarLander-v2 --exp\_name q2\_doubledqn\_1 --double\_q --seed 1

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name LunarLander-v2 --exp\_name q2\_doubledqn\_2 --double\_q --seed 2

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name LunarLander-v2 --exp\_name q2\_doubledqn\_3 --double\_q --seed 3

生成run-log文件后可视化遇到问题，未解决

1. 研究Q-learning对超参数的稳健性，命令行执行如下代码：

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name PongNoFrameskip-v4 --exp\_name q3\_hparam1 --batch\_size 16

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name PongNoFrameskip-v4 --exp\_name q3\_hparam2 --batch\_size 64

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_dqn.py" --env\_name PongNoFrameskip-v4 --exp\_name q3\_hparam3 --batch\_size 128

生成run-log文件后可视化遇到问题，未解决

1. 用actor-critic解CartPole-v0环境，命令行执行如下代码：

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_actor\_critic.py" --env\_name CartPole-v0 -n 100 -b 1000 --exp\_name 1\_1 -ntu 1 -ngsptu 1

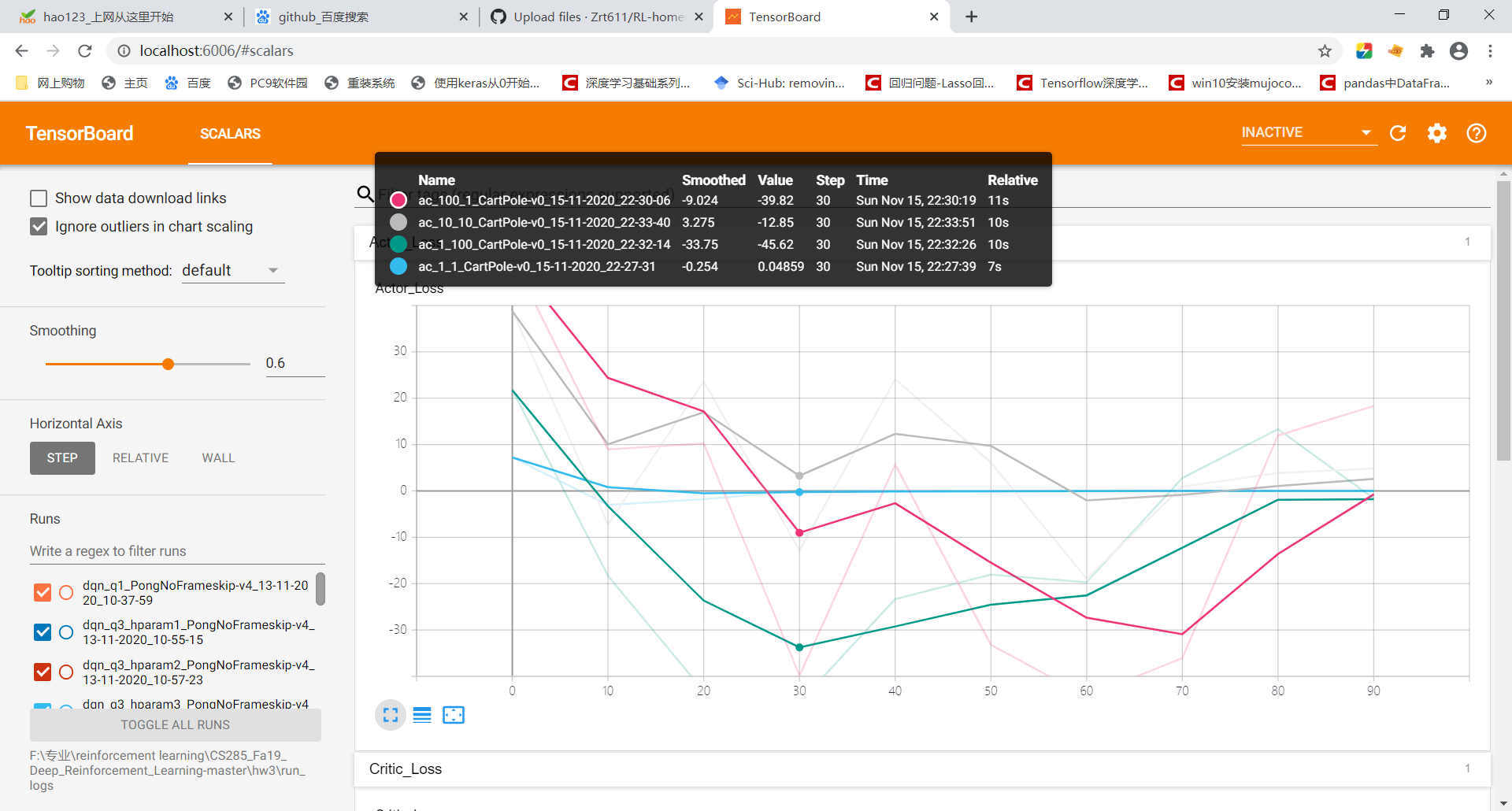
python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_actor\_critic.py" --env\_name CartPole-v0 -n 100 -b 1000 --exp\_name 100\_1 -ntu 100 -ngsptu 1

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_actor\_critic.py" --env\_name CartPole-v0 -n 100 -b 1000 --exp\_name 1\_100 -ntu 1 -ngsptu 100

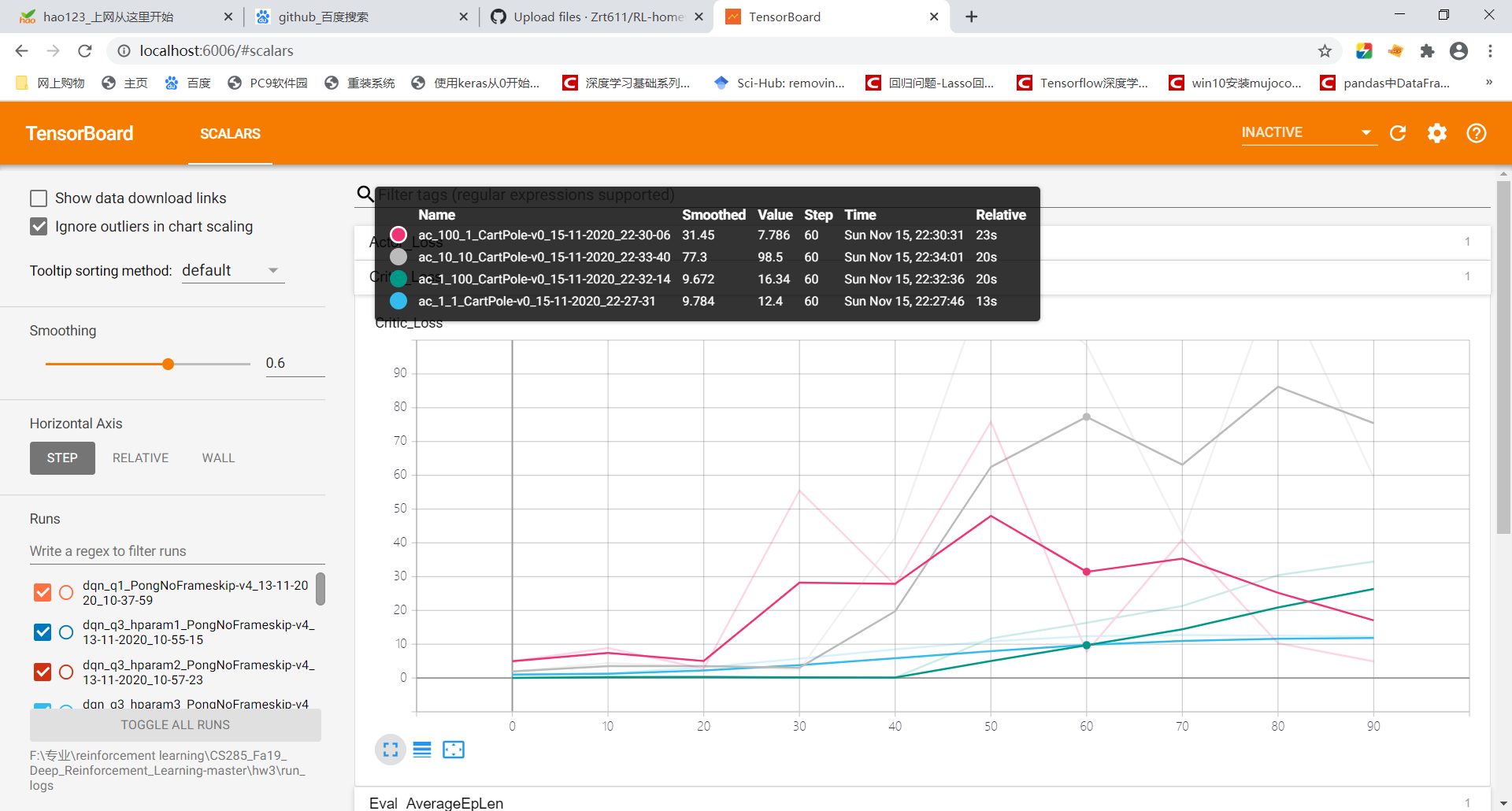
python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_actor\_critic.py" --env\_name CartPole-v0 -n 100 -b 1000 --exp\_name 10\_10 -ntu 10 -ngsptu 10

可视化结果如图：

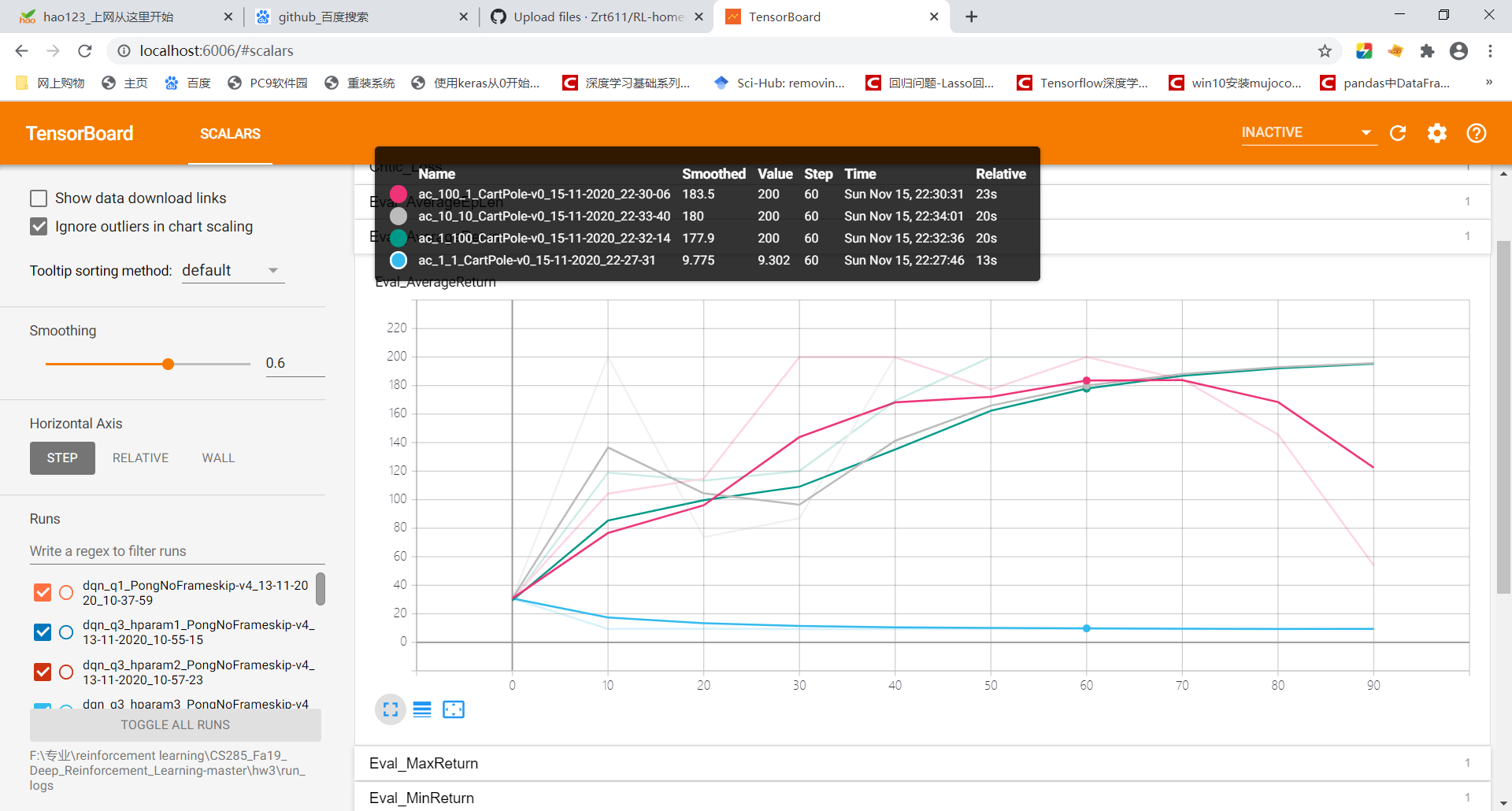
Actor-loss如图所示：



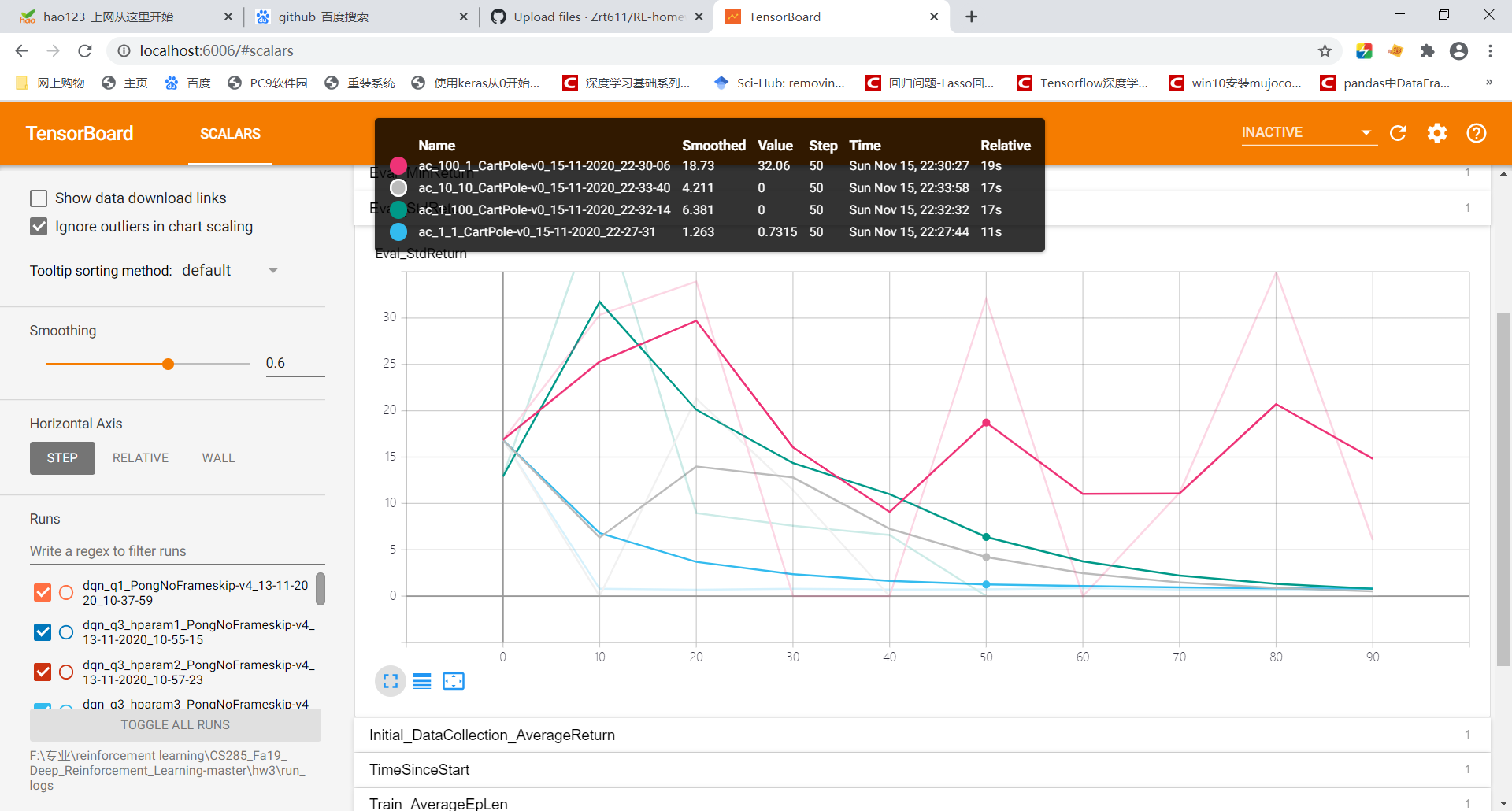
Critic-loss如图所示：



平均回报：



方差：



1. 综合对比average return和方差最优的超参数组合是10-10，用选取出来的最优超参数组合去解InvertedPendulum 和 HalfCheetah环境，命令行执行如下代码：

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_actor\_critic.py" --env\_name InvertedPendulum-v2 --ep\_len 1000 --discount 0.95 -n 100 -l 2 -s 64 -b 5000 -lr 0.01 --exp\_name 10\_10 -ntu 10 -ngsptu 10

python "F:\专业\reinforcement learning\CS285\_Fa19\_Deep\_Reinforcement\_Learning-master\hw3\cs285\scripts\run\_hw3\_actor\_critic.py" --env\_name HalfCheetah-v2--ep\_len 150 --discount 0.90 --scalar\_log\_freq 1 -n 150 -l 2 -s 32 -b 30000 -eb 1500 -lr 0.02 --exp\_name 10\_10 -ntu 10 -ngsptu 10

license到期了，未能生成run-log：

