**SB3 Default:**

class stable\_baselines.ppo2.PPO2 (policy, env,

gamma=0.99,

n\_steps=128,

ent\_coef=0.01,

learning\_rate=0.00025,

vf\_coef=0.5,

max\_grad\_norm=0.5,

lam=0.95,

nminibatches=4, noptepochs=4, cliprange=0.2, cliprange\_vf=None, verbose=0, tensorboard\_log=None, \_init\_setup\_model=True, policy\_kwargs=None, full\_tensorboard\_log=False, seed=None, n\_cpu\_tf\_sess=None)

**Hardmaru:**

imesteps\_per\_actorbatch=4096,

clip\_param=0.2,

entcoeff=0.0,

optim\_epochs=10,

optim\_stepsize=3e-4,

optim\_batchsize=64,

gamma=0.99,

lam=0.95,

schedule='linear',

verbose=2

**PPO template:**



SB3 default랑 hardmaru 가 다른 hyperparameter만 고려

(def): SB3 default 라는 뜻

n\_steps: 4096, 2048

batch\_size: 64, 128

learning\_rate: lin(0.00025), lin(0.0003), lin(0.0007), 0.00025(def), 0.0003, 0.0007, ,

verbose: 2, 1, 0(def)

ent\_coef: 0.0(def), 0.1

n\_epochs =10 (fixed)

**Standard (hardmaru):**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**진성:**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**n\_steps: 2048**

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

n\_steps: 4096

**batch\_size: 128**

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

n\_steps: 4096

batch\_size: 64

**learning\_rate: lin(0.0003)**

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**정엽:**

n\_steps: 4096

batch\_size: 64

**learning\_rate: lin(0.0007)**

verbose: 2

ent\_coef: 0.0

n\_epochs =10

n\_steps: 4096

batch\_size: 64

**learning\_rate: 0.00025**

verbose: 2

ent\_coef: 0.0

n\_epochs =10

n\_steps: 4096

batch\_size: 64

**learning\_rate: 0.0003**

verbose: 2

ent\_coef: 0.0

n\_epochs =10

n\_steps: 4096

batch\_size: 64

**learning\_rate: 0.0007**

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**이솔:**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

**verbose: 1**

ent\_coef: 0.0

n\_epochs =10

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

**verbose: 0**

ent\_coef: 0.0

n\_epochs =10

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

**ent\_coef: 0.1**

n\_epochs =10

Learning rate:

(lin(0.00025)=lin(0.0007)=0.00025)>=lin(0.0003)>0.0007>0.0003

N\_steps:

4096>2048

Batch size:

64>128

Ent\_coef:

0.0>0.1

**추가로 해볼 것**

Gamma: 0.95, 0.99(def), 0.999

Clip range: 0.1, 0.2(def)

Vf\_coef: 0, 0.5(def), 1

gae \_rambda: 0.92, 0.95(def), 0.98

N\_epoch: 5, 10(def), 20

**진성:**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**Gamma: 0.95**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**Gamma: 0.999**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**Clip range: 0.1**

**정엽:**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**Vf\_coef: 0**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**Vf\_coef: 1**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**Gae\_rambda: 0.92**

**이솔:**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

**Gae\_rambda: 0.98**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

**n\_epochs =5**

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

**n\_epochs =20**

Learning rate:

(lin(0.00025)=lin(0.0007)=0.00025)>=lin(0.0003)>0.0007>0.0003

N\_steps:

4096>2048

Batch size:

64>128

Ent\_coef:

0.0>0.1

Gamma:

0.99>0.95>0.999

Clip range:

0.2>0.1

N\_epoch:

20>=10>5

Gae\_rambda:

0.95>0.98

Hyperparameter 조합

이솔

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

정엽

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.0007)

verbose: 2

ent\_coef: 0.0

n\_epochs =10

진성

n\_steps: 4096

batch\_size: 64

learning\_rate: lin(0.00025)

verbose: 2

ent\_coef: 0.0

n\_epochs =20

상대 가능한 것:

지금까지 우리가 training 시킨 애들(poorly trained) A2C, PPO

* PPO: js\_trial\_6
* A2C: jowrap\_js\_trial\_1

Random agent

Default ai

Hardmaru1

Hardmaru2

Hardmaru3

훈련방법

학습시킨 애 계속 loading해서 상대바꾸면서 진행

1. SlimeVolleyOpponentEnv class 사용하여 계속 돌리기

5e7\*10

2. poorly -random-default 순서

((Poorly ppo 3e7)+(poorly a2c 3e7)+(random 1e7)

+(default 9e7)+(hardmaru1 3e7)+(hardmaru2 3e7)+(hardmaru3 3e7))\*2