



Reinforcement Learning

나는 강화학습으로 축구한다

Google Research와 Manchester City F.C.의
인공지능 축구 프로젝트를 대한민국에서
재조명하고 직접 구현해보는 캠프



SESSION

12

Copyright Notice

These slides are available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite the author as the source of the slides.



오후에 끝내는 실전 RL Techniques

- Log Analysis
- Tensorboard
- Steps
- Custom Reward Function
- Vector Normalization
- Neural Network Variation
- Reward Shaping
- Parallel Training
- Hyperparameter Tuning

Hands-On Exercise

Custom Reward Function

LAB
07



Live



Hands-On Exercise

Training Steps

LAB
08



Live



How Many Steps Are Sufficient?

```
import gfootball.env as football_env
from stable_baselines3 import PPO

env = football_env.create_environment(
    env_name="academy_empty_goal_close",
    representation="simple115v2"
)

env.reset()

model = PPO("MlpPolicy", env, verbose=1)

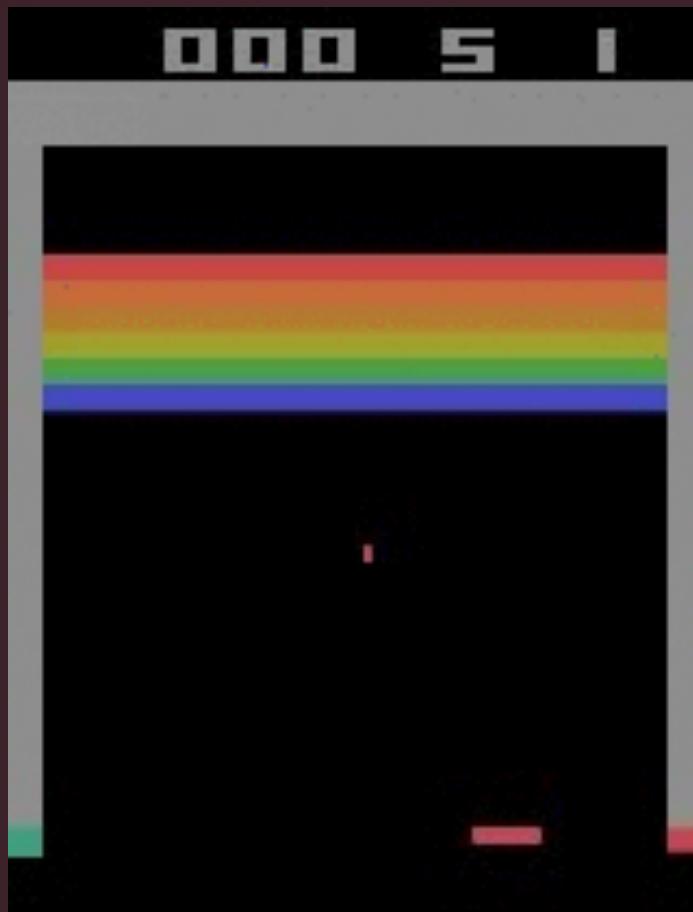
model.learn(total_timesteps=10_000)

model.save("model-ppo.zip")

env.close()
```



Steps



Initial



After 15 mins.



After 30 mins.

Hands-On Exercise

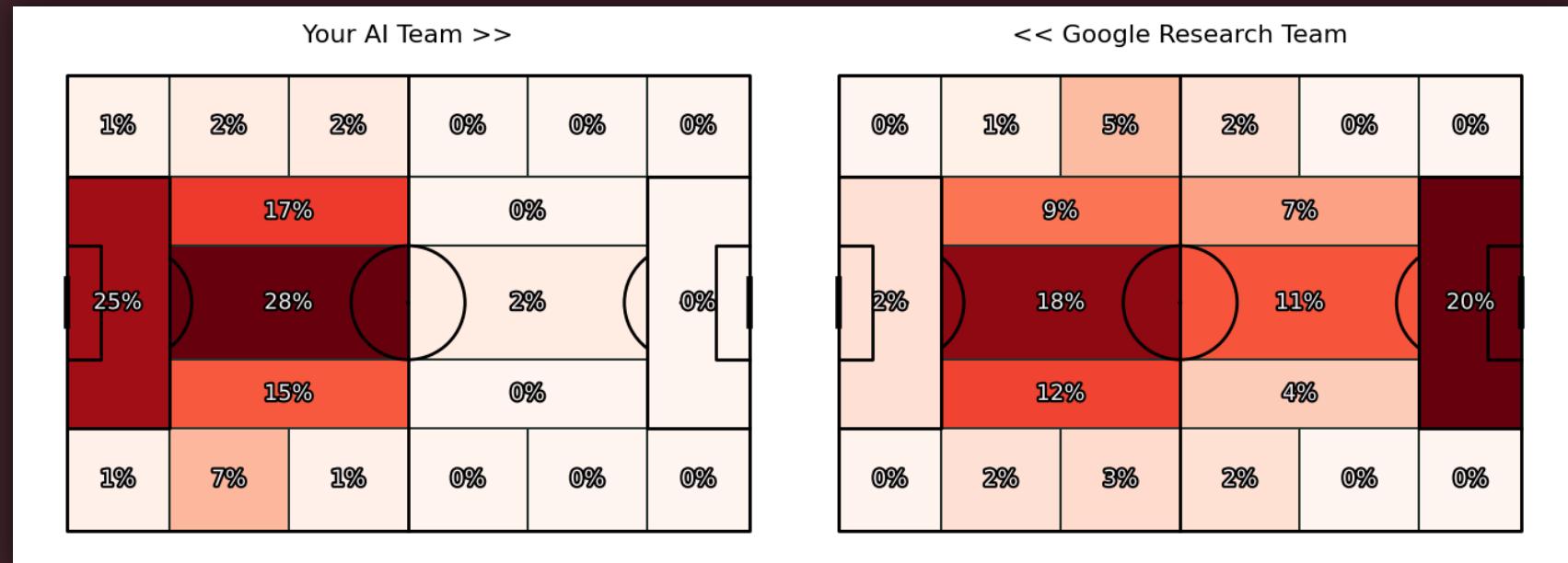
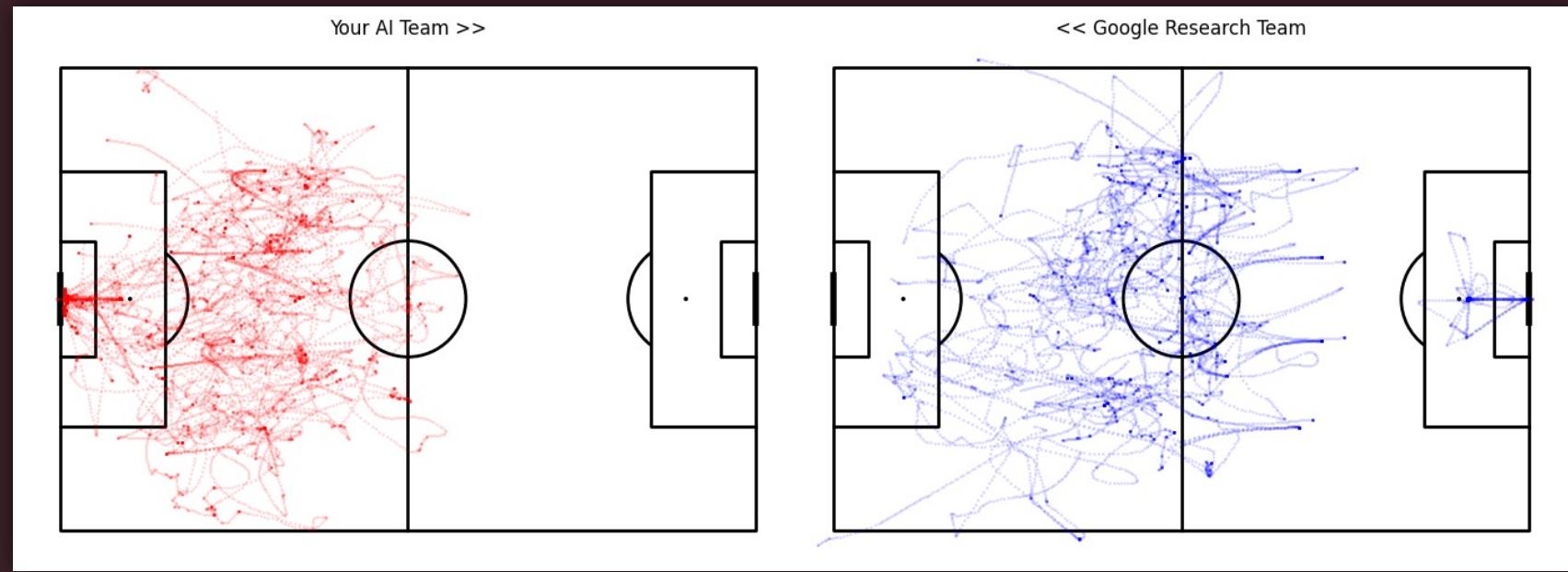
Log Generation and Analysis

LAB
09



Live





Hands-On Exercise

Monitoring via Tensorboard

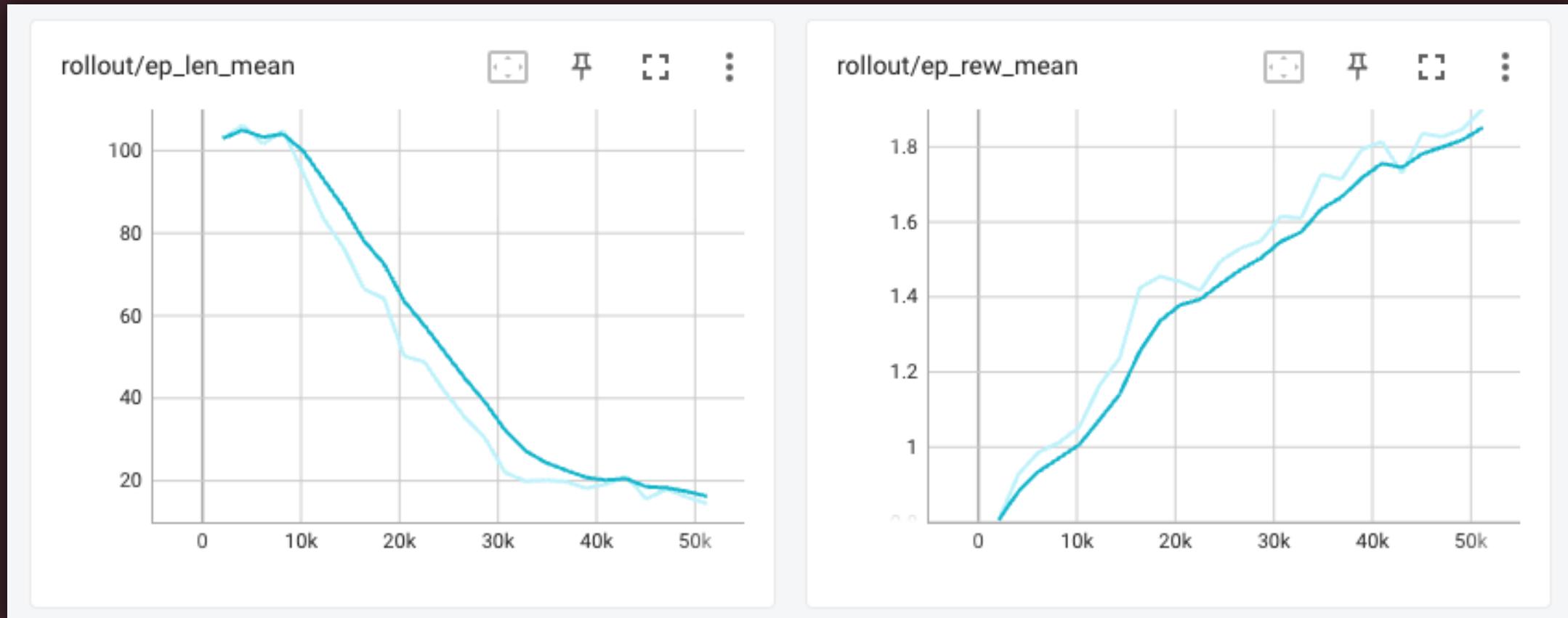
LAB
10



Live



Tensorboard Integration



Hands-On Exercise

Neural Network Variation

LAB
11

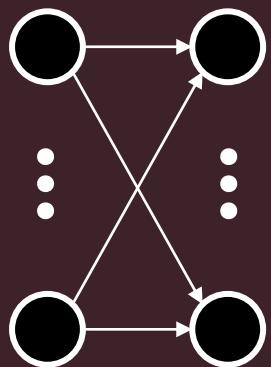


Live

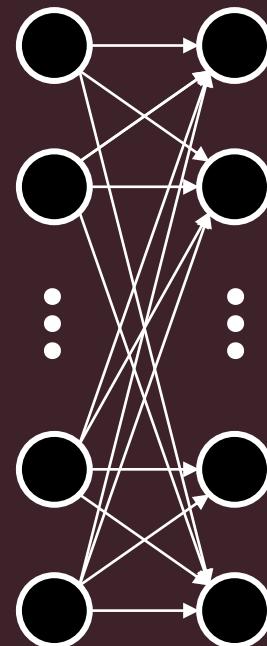


Neural Network Variation

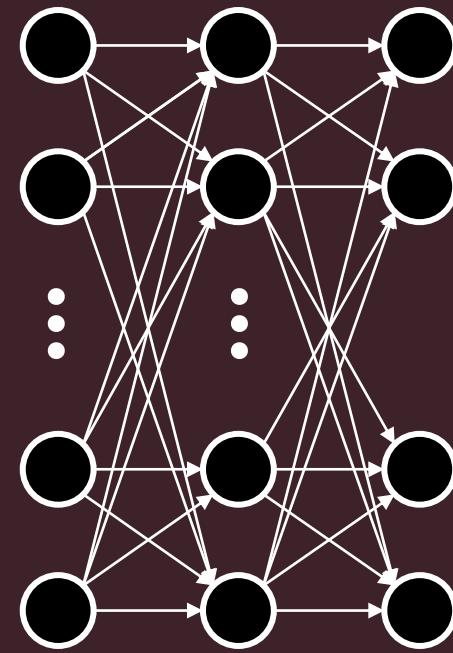
Two Layers
with 32 Nodes Each



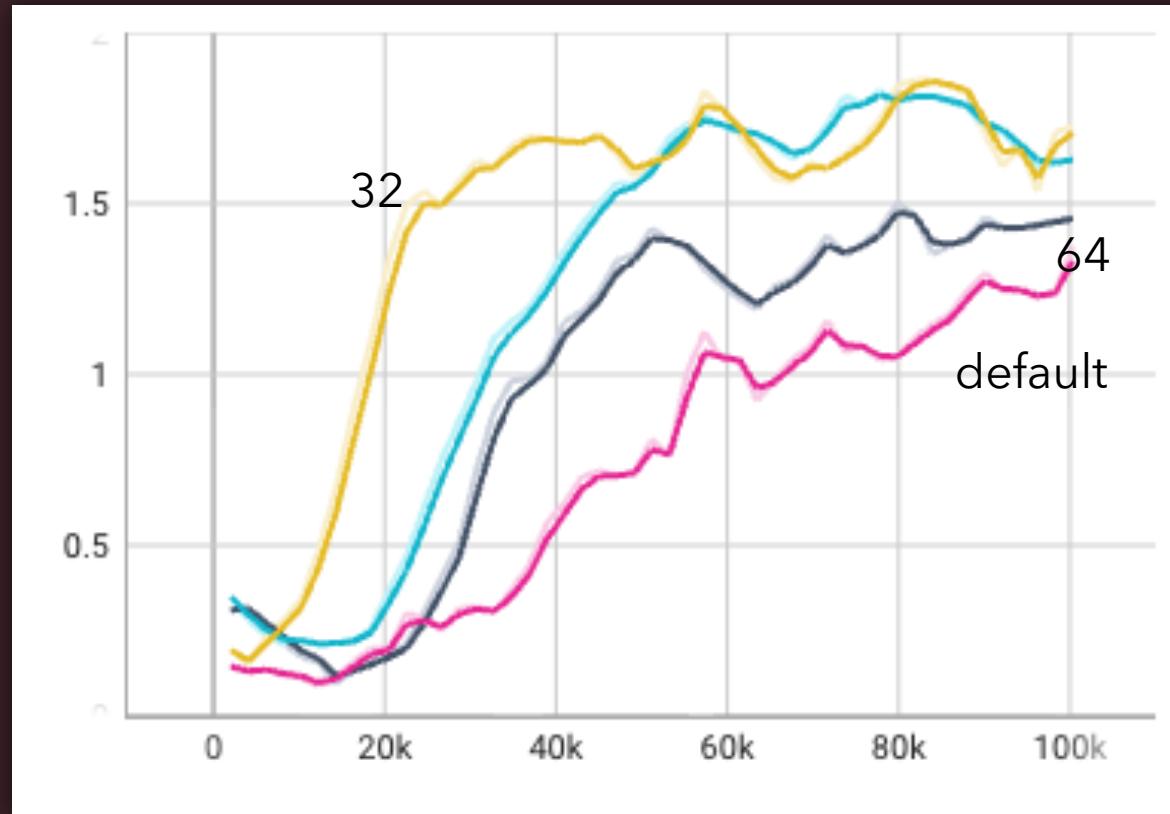
Two Layers
with 64 Nodes Each



Three Layers
with 128 Nodes Each



Neural Network Variation



- default
- 32-32
- 64-64
- 128-128

Hands-On Exercise

Checkpoints

LAB
13

LAB
14

LAB
15

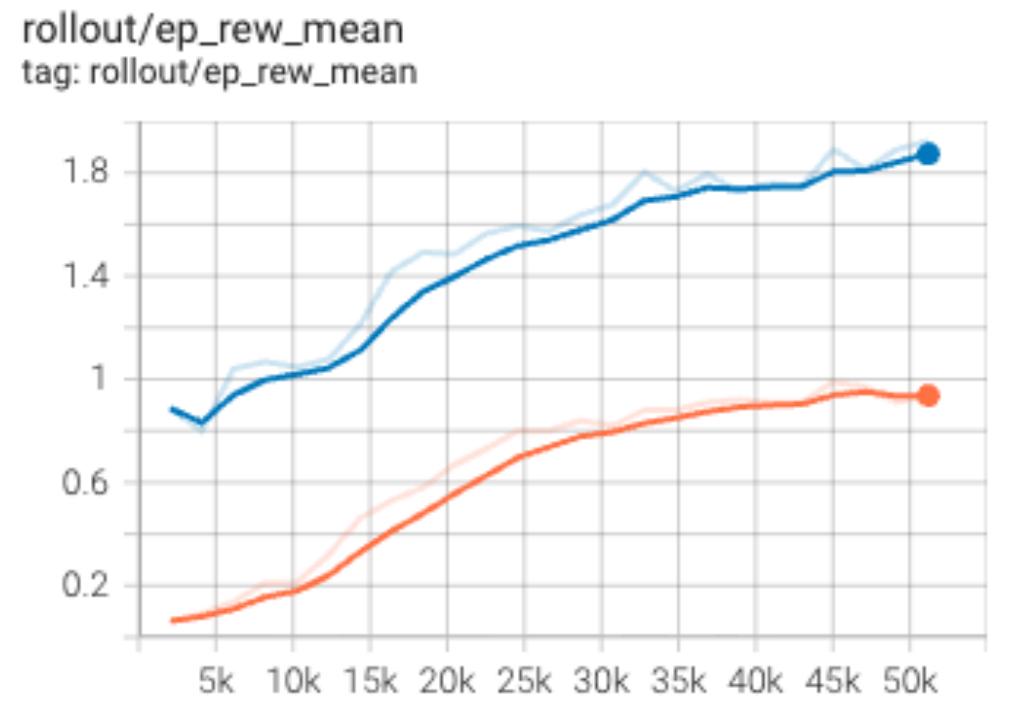
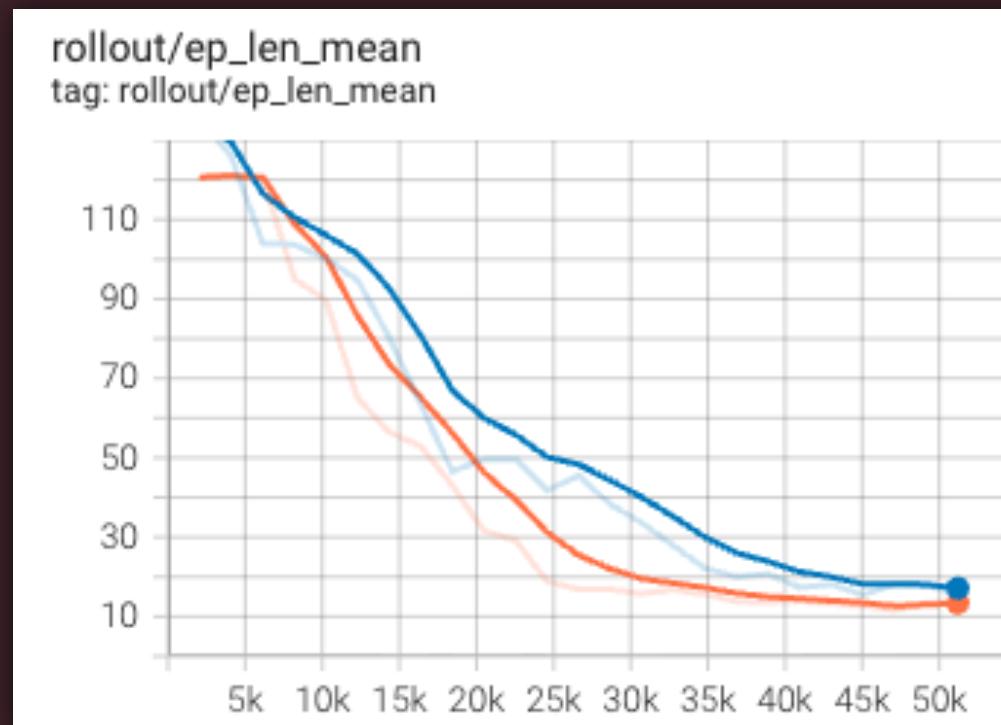


Live



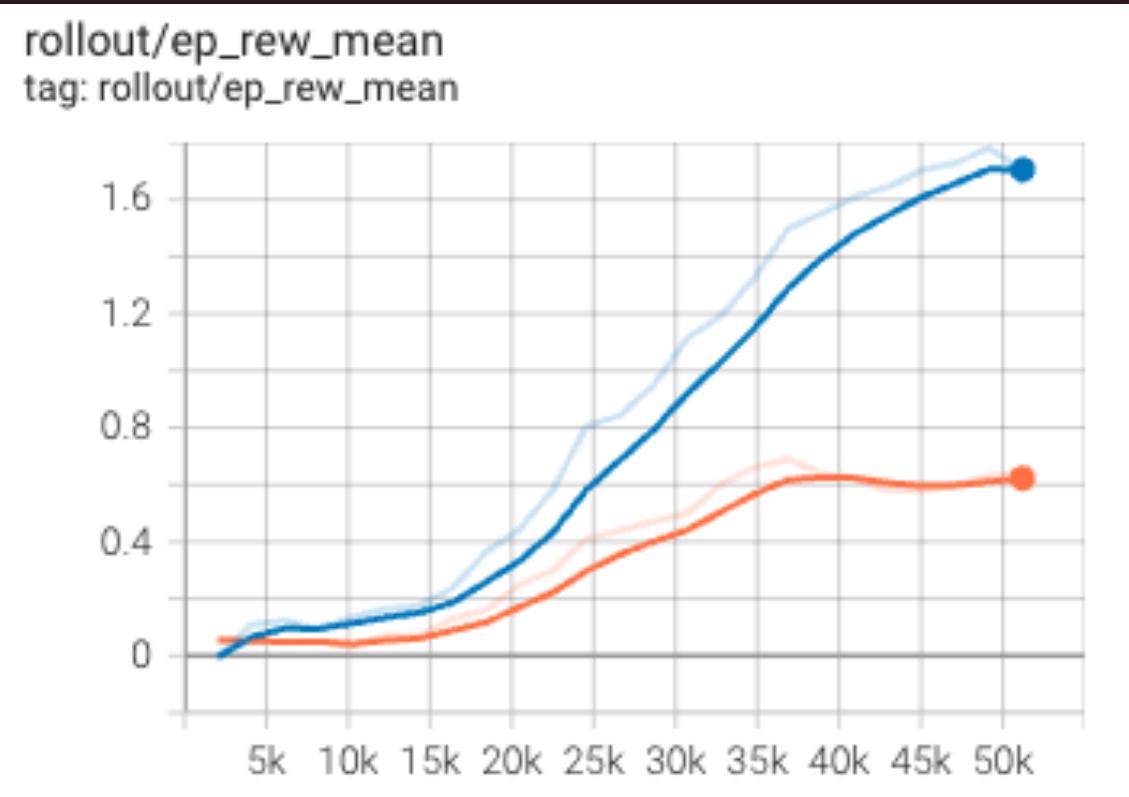
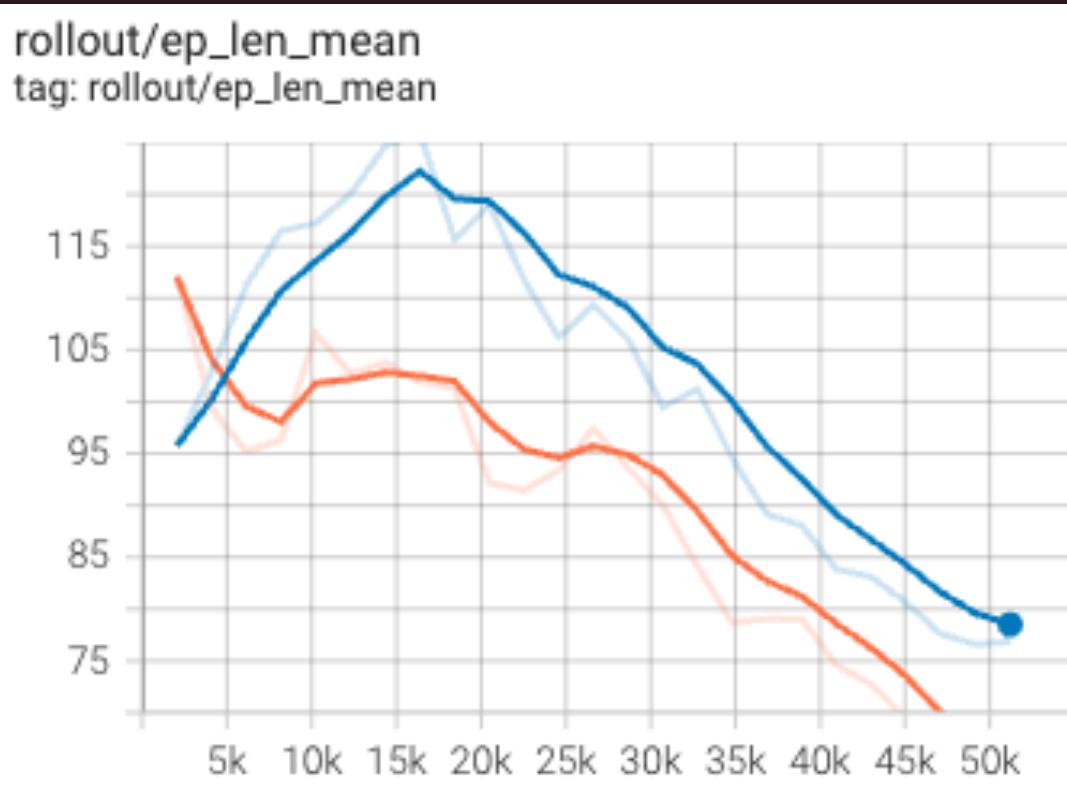
academy_empty_goal_close

with vs. without checkpoints



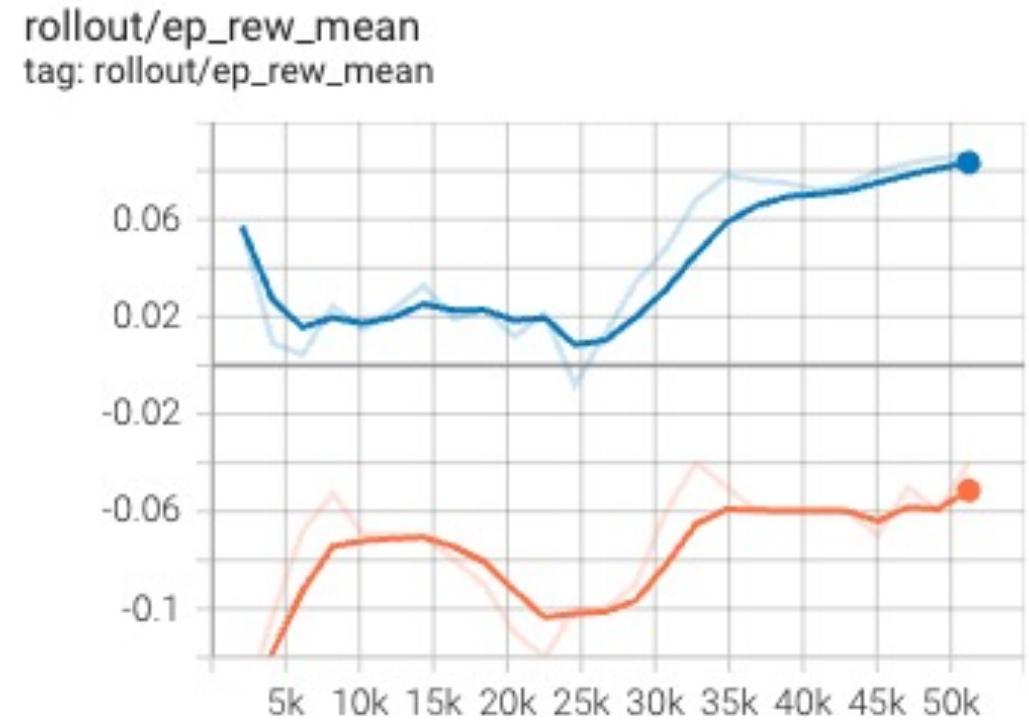
academy_empty_goal

with vs. without checkpoints



academy_run_to_score_with_keeper

PPO scoring vs. scoring + checkpoints



Hands-On Exercise

Vector Normalization

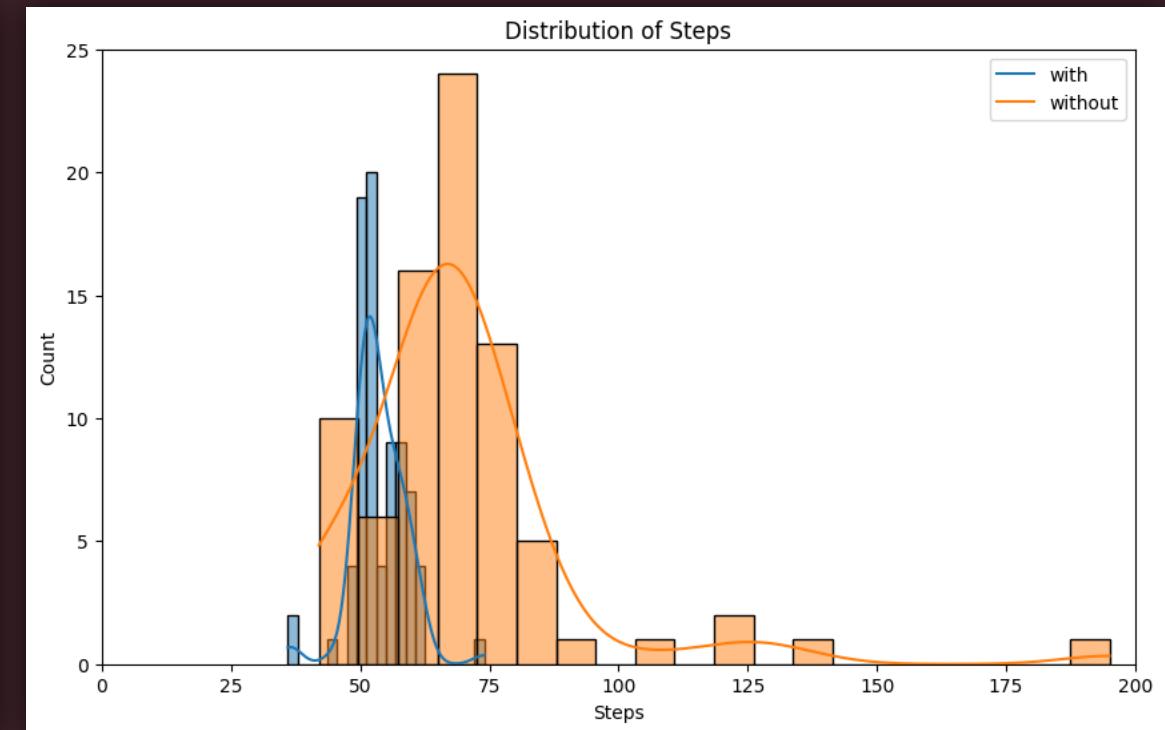
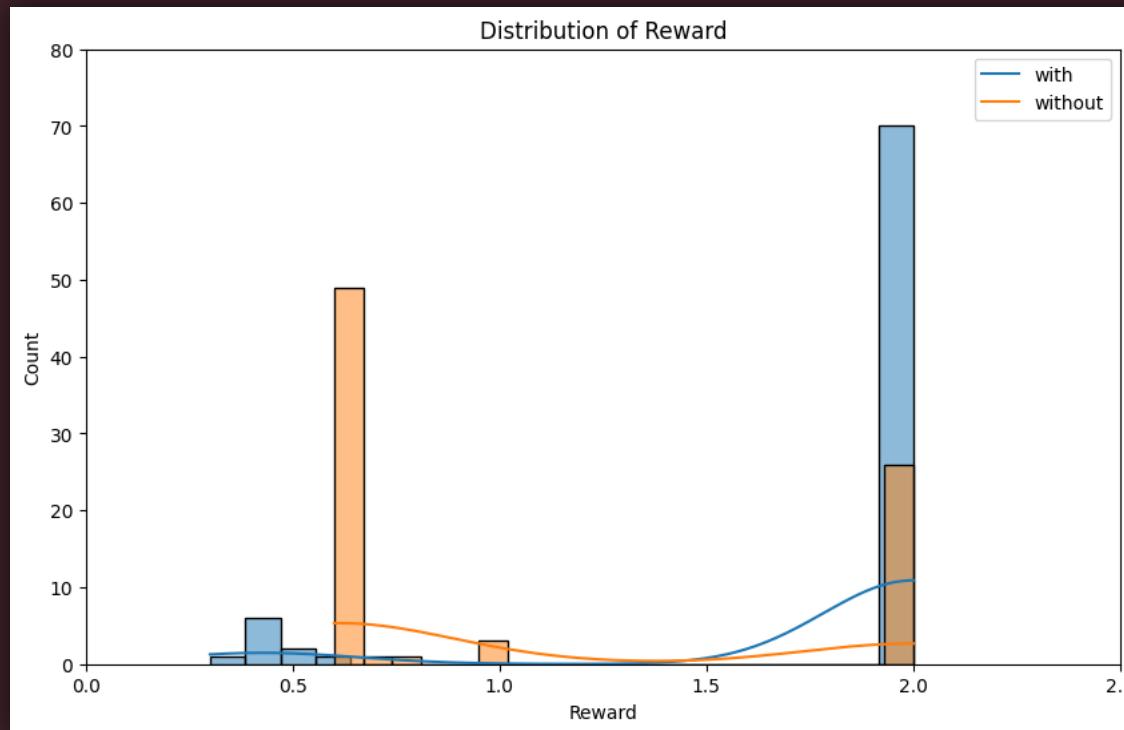
LAB
16



Live

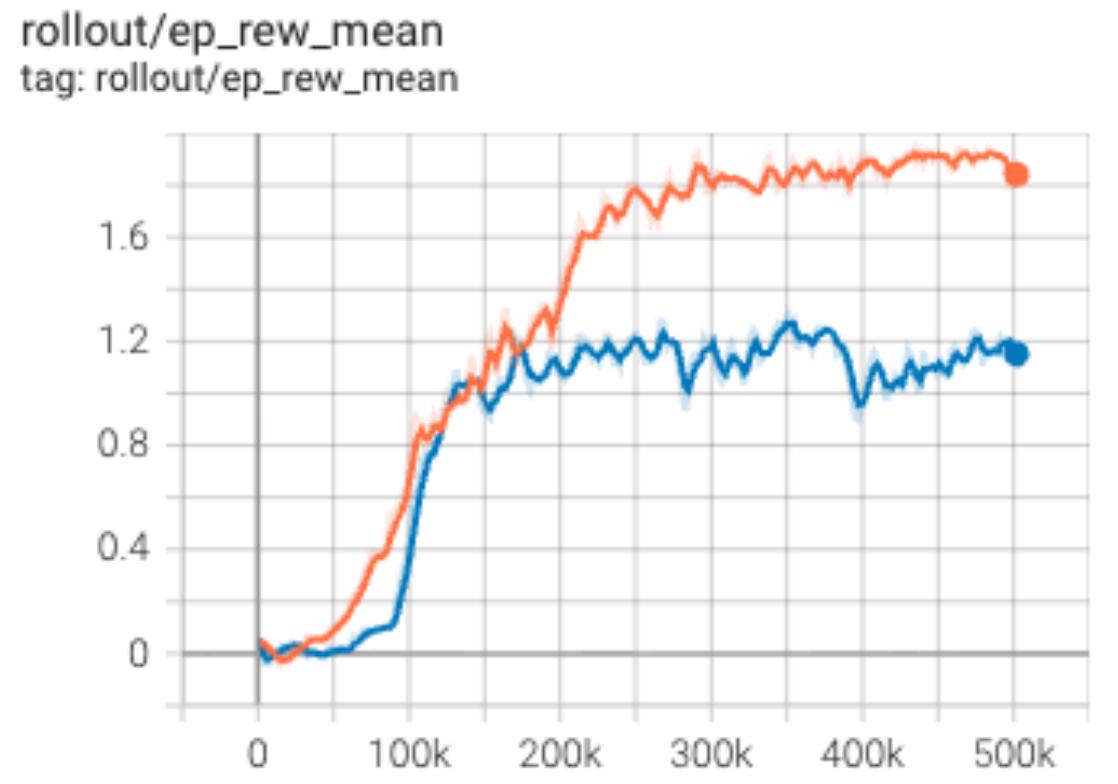
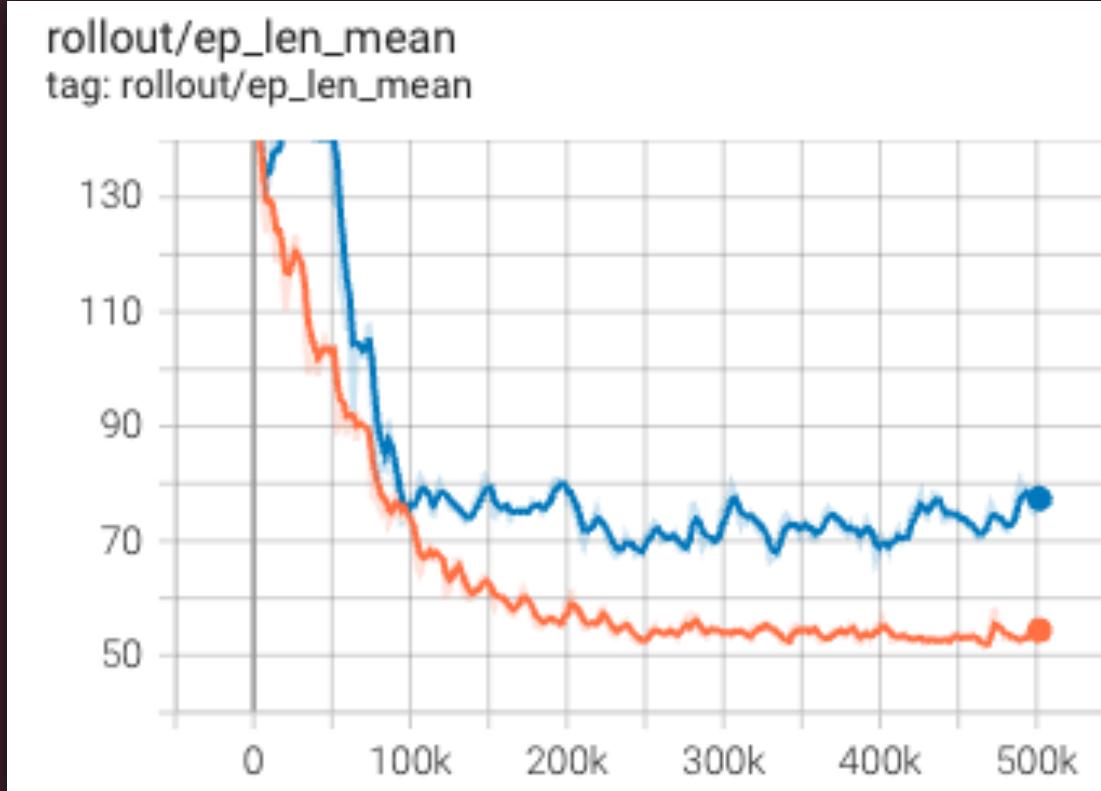


Raw vs. Vector Normalization



academy_run_to_score_with_keeper

PPO + scoring + checkpoints + vecnorm vs. no vecnorm



Hands-On Exercise

Parallel Training

LAB
17



Live



Parallel Training



Single Environment

Multiple Environments
(n_envs = 4)



Hands-On Exercise

Hyperparameter Optimization

LAB
18

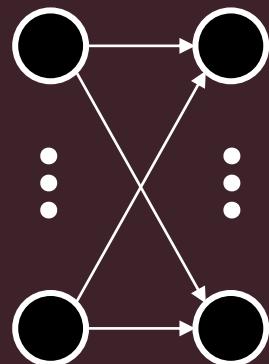


Live



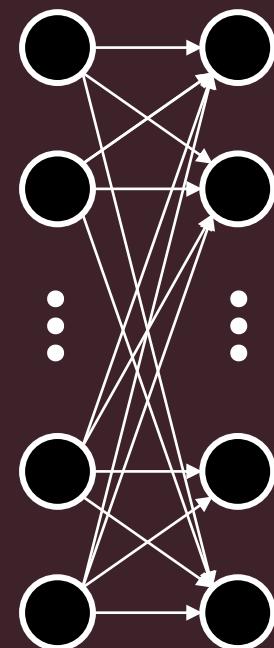
Which one is better? Which combination is better?

Two Layers
with 32 Nodes Each

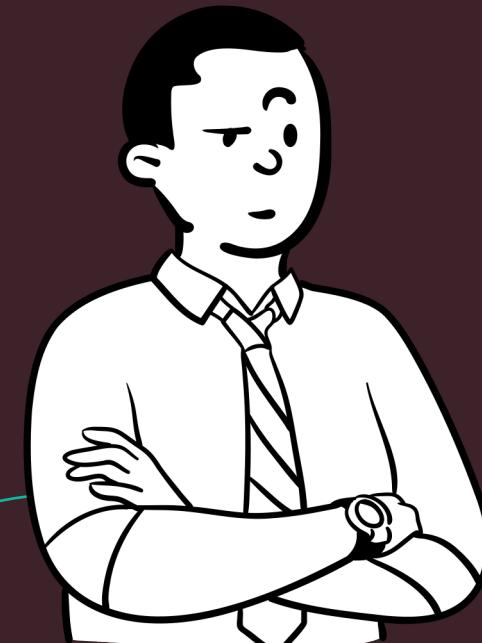


vs.

Two Layers
with 64 Nodes Each



0.1 vs. 0.2 for progress reward?



Optuna

```
[I 2026-01-20 09:45:59,626] Trial 0 finished with value: 0.10000000149011612 and parameters:  
{'learning_rate': 0.00010115920670134792, 'net_arch': 'large', 'n_steps': 2048, 'batch_size': 64, 'n_epochs': 14}.  
Best is trial 0 with value: 0.10000000149011612.
```

```
[I 2026-01-20 09:47:07,549] Trial 1 finished with value: 0.10000000149011612 and parameters:  
{'learning_rate': 0.0004619308809930084, 'net_arch': 'medium', 'n_steps': 2048, 'batch_size': 256, 'n_epochs': 13}.  
Best is trial 0 with value: 0.10000000149011612.
```

```
[I 2026-01-20 09:48:12,517] Trial 2 finished with value: 0.10000000149011612 and parameters:  
{'learning_rate': 0.0005108058672138399, 'net_arch': 'large', 'n_steps': 1024, 'batch_size': 64, 'n_epochs': 12}.  
Best is trial 0 with value: 0.10000000149011612.
```

```
100%| #####| 3/3 [03:29<00:00, 69.96s/it]
```

Mean Reward: 0.10000000149011612

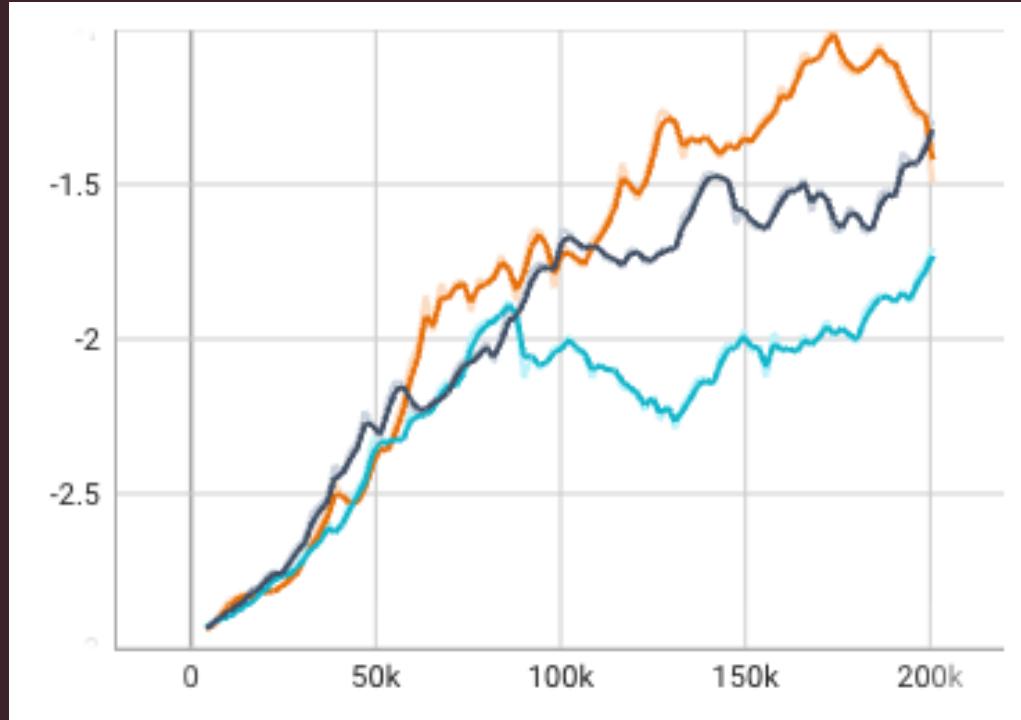
Best Hyperparameters:

```
learning_rate: 0.00010115920670134792  
net_arch: large  
n_steps: 2048  
batch_size: 64  
n_epochs: 14
```

A close-up photograph of a person's hands working on a piece of wood. One hand holds a wooden chisel, and the other holds a metal ruler, marking a measurement on the wood. The background is blurred, showing more of the workshop environment.

Other Performance Metrics

train/entropy_loss



High Entropy: More Exploration
Low Entropy: More Deterministic

entropy_loss	Interpretation
-2.0	Very stochastic policy
-0.5	Moderate exploration
-0.05	Near-deterministic
≈ 0	Almost no exploration

Early Training

- High entropy (large negative entropy_loss)

Mid Training

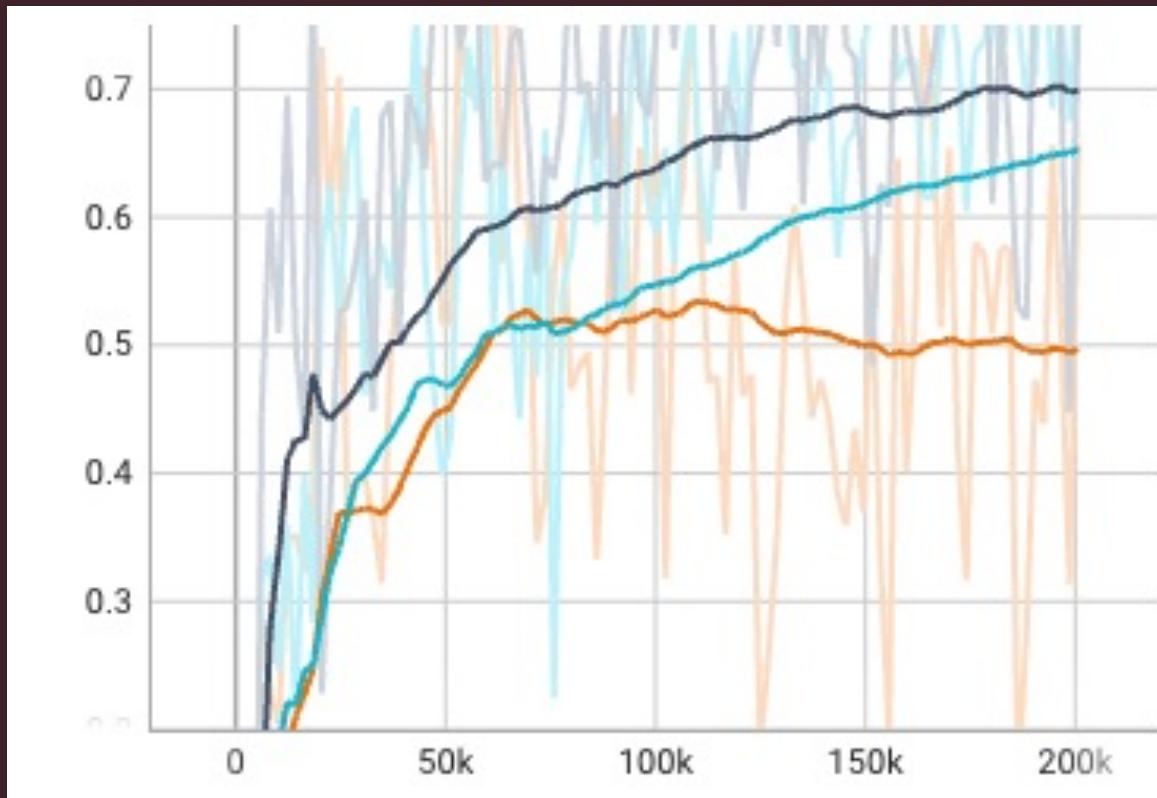
- Gradual increase toward zero

Late Training

- Small negative value
- Policy mostly deterministic

19 discrete actions \rightarrow max entropy $\approx \log(19) \approx 2.94$

train/explained_variance



Interpretation

- $\approx 1.0 \rightarrow$ value function fits well
- $\approx 0.0 \rightarrow$ no better than predicting mean
- $< 0 \rightarrow$ very bad value predictions

Paper

The screenshot shows a web browser window with a purple header bar. The address bar contains the URL arxiv.org/abs/1907.11180. The page itself is a Cornell University-hosted arXiv document. At the top left is the Cornell University logo. To its right, a message says "We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors." with a "Donate" button. The arXiv navigation bar includes a search bar with dropdown menus for "Search..." and "All fields", and links for "Help | Advanced Search". The main content area has a red header "Computer Science > Machine Learning". Below it, a note says "[Submitted on 25 Jul 2019 (v1), last revised 14 Apr 2020 (this version, v2)]". The title "Google Research Football: A Novel Reinforcement Learning Environment" is prominently displayed. Below the title is a list of authors: Karol Kurach, Anton Raichuk, Piotr Stańczyk, Michał Zajęc, Olivier Bachem, Lasse Espeholt, Carlos Riquelme, Damien Vincent, Marcin Michalski, Olivier Bousquet, Sylvain Gelly. A detailed abstract follows, describing the creation of a physics-based 3D simulator for reinforcement learning. On the right side of the page, there is a sidebar titled "Access Paper:" which includes links to "View PDF", "TeX Source", "Other Formats", and a "view license" link. It also shows the current browse context as "cs.LG" and provides links for "Change to browse by:" and "References & Citations" to various databases like NASA ADS, Google Scholar, and Semantic Scholar. At the bottom of the sidebar, there is a link to "1 blog link" and "DBLP – CS Bibliography".