**TRAINING PHASE**

LR usando PPO = 3e-4, LR usando TRPO = 1e-3 (se non specificati)

|  |  |  |
| --- | --- | --- |
| AGENT | TIMESTEPS | RESULT (mean +/- std dev on the reward) |
| PPO | 1K | 75.59 +/- 1.54 |
| PPO | 10K | 74.05 +/- 1.41 |
| PPO | 20K | 221.78 +/- 1.13 |
| PPO | 50K | 366.84 +/- 1.48 |
| PPO | 100K | 575.52 +/- 33.32 |
| TRPO | 1K | 67.73 +/- 1.52 |
| TRPO | 10K | 253.03 +/- 1.57 |
| TRPO | 20K | 200.66 +/- 0.75 |
| TRPO | 50K | 388.70 +/- 218.39 |
| TRPO | 100K | 827.65 +/- 260.86 |
| TRPO (lr=3e-4) | 50K | 289.30 +/- 1.68 |
| TRPO (lr=3e-4) | 100K | 404.18 +/- 3.46 |
| TRPO (lr = 3e-4) | 10K | 290.37 +/- 1.97 |
| PPO (lr = 3e-5) | 100K | 181.55 +/- 2.18 |
| PPO (lr = 1e-3) | 100K | 1369.29 +/- 123.25 |
|  |  |  |
|  |  |  |
|  |  |  |

**TABLE FOR STEP 2.3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ALGORITHM | HYPERPARAMETERS | SOURCE-SOURCE AVERAGE RETURN | SOURCE-TARGET AVERAGE RETURN | TARGET-TARGET AVERAGE RETURN |
| PPO | LR=3E-4, TS=100K | 1.85 | 1.74 |  |
| PPO | LR=3E-4, TS=50K | 1.85 | 1.73 |  |
| PPO | LR=3E-4, TS=20K | 1.74 | 1.65 |  |
| PPO | LR=3E-5, TS=100K | 1.92 | 1.78 |  |
| PPO | LR=1E-3, TS=100K | 1.83 | 1.74 |  |
|  |  |  |  |  |
|  |  |  |  |  |