**train\_selftouch description**

Key code description

Intrinsic reward mechanism:

python

active\_sensors = np.sum(obs['touch'] > 1e-6)

return active\_sensors / self.max\_touch\_sensors

Calculate the proportion of activated touch sensors as intrinsic rewards to encourage MIMo to explore more physical contact

Environment configuration:

Use examples/config\_selftouch.yml to pre-configure the environment

Enable tactile sensors, disable vision and vestibular systems

Use spring damper to drive the model

Reinforcement learning model:

Use PPO algorithm (Proximal Policy Optimization)

The policy network automatically adapts to the multi-input observation space

Key parameters optimize the infant learning scenario

输出

(babybench) C:\Users\Administrator\BabyBench2025\_Starter\_Kit>python train\_selftouch.py

Using cpu device

Wrapping the env with a `Monitor` wrapper

Wrapping the env in a DummyVecEnv.

---------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.43 |

| time/ | |

| fps | 125 |

| iterations | 1 |

| time\_elapsed | 16 |

| total\_timesteps | 2048 |

---------------------------------

-----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.47 |

| time/ | |

| fps | 81 |

| iterations | 2 |

| time\_elapsed | 50 |

| total\_timesteps | 4096 |

| train/ | |

| approx\_kl | 0.072494775 |

| clip\_fraction | 0.462 |

| clip\_range | 0.2 |

| entropy\_loss | -42.6 |

| explained\_variance | -7.3668966 |

| learning\_rate | 0.0003 |

| loss | -0.57 |

| n\_updates | 10 |

| policy\_gradient\_loss | -0.0994 |

| std | 1 |

| value\_loss | 0.0193 |

-----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.57 |

| time/ | |

| fps | 78 |

| iterations | 3 |

| time\_elapsed | 77 |

| total\_timesteps | 6144 |

| train/ | |

| approx\_kl | 0.08659227 |

| clip\_fraction | 0.559 |

| clip\_range | 0.2 |

| entropy\_loss | -42.8 |

| explained\_variance | -9.435876 |

| learning\_rate | 0.0003 |

| loss | -0.57 |

| n\_updates | 20 |

| policy\_gradient\_loss | -0.11 |

| std | 1.01 |

| value\_loss | 0.00516 |

----------------------------------------

-----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.64 |

| time/ | |

| fps | 74 |

| iterations | 4 |

| time\_elapsed | 110 |

| total\_timesteps | 8192 |

| train/ | |

| approx\_kl | 0.088341534 |

| clip\_fraction | 0.567 |

| clip\_range | 0.2 |

| entropy\_loss | -43 |

| explained\_variance | -4.982903 |

| learning\_rate | 0.0003 |

| loss | -0.559 |

| n\_updates | 30 |

| policy\_gradient\_loss | -0.111 |

| std | 1.02 |

| value\_loss | 0.00275 |

-----------------------------------------

-----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.73 |

| time/ | |

| fps | 71 |

| iterations | 5 |

| time\_elapsed | 142 |

| total\_timesteps | 10240 |

| train/ | |

| approx\_kl | 0.096172415 |

| clip\_fraction | 0.583 |

| clip\_range | 0.2 |

| entropy\_loss | -43.2 |

| explained\_variance | -6.4073896 |

| learning\_rate | 0.0003 |

| loss | -0.581 |

| n\_updates | 40 |

| policy\_gradient\_loss | -0.112 |

| std | 1.03 |

| value\_loss | 0.00186 |

-----------------------------------------

-----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.74 |

| time/ | |

| fps | 71 |

| iterations | 6 |

| time\_elapsed | 171 |

| total\_timesteps | 12288 |

| train/ | |

| approx\_kl | 0.112526834 |

| clip\_fraction | 0.604 |

| clip\_range | 0.2 |

| entropy\_loss | -43.5 |

| explained\_variance | -2.9554896 |

| learning\_rate | 0.0003 |

| loss | -0.59 |

| n\_updates | 50 |

| policy\_gradient\_loss | -0.113 |

| std | 1.03 |

| value\_loss | 0.00133 |

-----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.79 |

| time/ | |

| fps | 70 |

| iterations | 7 |

| time\_elapsed | 203 |

| total\_timesteps | 14336 |

| train/ | |

| approx\_kl | 0.11635405 |

| clip\_fraction | 0.622 |

| clip\_range | 0.2 |

| entropy\_loss | -43.7 |

| explained\_variance | -2.45538 |

| learning\_rate | 0.0003 |

| loss | -0.584 |

| n\_updates | 60 |

| policy\_gradient\_loss | -0.115 |

| std | 1.04 |

| value\_loss | 0.000939 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.84 |

| time/ | |

| fps | 71 |

| iterations | 8 |

| time\_elapsed | 230 |

| total\_timesteps | 16384 |

| train/ | |

| approx\_kl | 0.13436627 |

| clip\_fraction | 0.626 |

| clip\_range | 0.2 |

| entropy\_loss | -44 |

| explained\_variance | -1.1417286 |

| learning\_rate | 0.0003 |

| loss | -0.582 |

| n\_updates | 70 |

| policy\_gradient\_loss | -0.115 |

| std | 1.05 |

| value\_loss | 0.000788 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.91 |

| time/ | |

| fps | 71 |

| iterations | 9 |

| time\_elapsed | 256 |

| total\_timesteps | 18432 |

| train/ | |

| approx\_kl | 0.13603157 |

| clip\_fraction | 0.641 |

| clip\_range | 0.2 |

| entropy\_loss | -44.2 |

| explained\_variance | -2.107596 |

| learning\_rate | 0.0003 |

| loss | -0.597 |

| n\_updates | 80 |

| policy\_gradient\_loss | -0.116 |

| std | 1.06 |

| value\_loss | 0.000732 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 5.99 |

| time/ | |

| fps | 72 |

| iterations | 10 |

| time\_elapsed | 281 |

| total\_timesteps | 20480 |

| train/ | |

| approx\_kl | 0.13637921 |

| clip\_fraction | 0.629 |

| clip\_range | 0.2 |

| entropy\_loss | -44.5 |

| explained\_variance | -1.0778098 |

| learning\_rate | 0.0003 |

| loss | -0.596 |

| n\_updates | 90 |

| policy\_gradient\_loss | -0.114 |

| std | 1.07 |

| value\_loss | 0.000588 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.05 |

| time/ | |

| fps | 73 |

| iterations | 11 |

| time\_elapsed | 305 |

| total\_timesteps | 22528 |

| train/ | |

| approx\_kl | 0.14304504 |

| clip\_fraction | 0.64 |

| clip\_range | 0.2 |

| entropy\_loss | -44.8 |

| explained\_variance | -1.0177405 |

| learning\_rate | 0.0003 |

| loss | -0.588 |

| n\_updates | 100 |

| policy\_gradient\_loss | -0.115 |

| std | 1.08 |

| value\_loss | 0.000509 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.12 |

| time/ | |

| fps | 72 |

| iterations | 12 |

| time\_elapsed | 341 |

| total\_timesteps | 24576 |

| train/ | |

| approx\_kl | 0.14819744 |

| clip\_fraction | 0.627 |

| clip\_range | 0.2 |

| entropy\_loss | -45 |

| explained\_variance | -0.5163778 |

| learning\_rate | 0.0003 |

| loss | -0.595 |

| n\_updates | 110 |

| policy\_gradient\_loss | -0.112 |

| std | 1.09 |

| value\_loss | 0.000386 |

----------------------------------------

-----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.2 |

| time/ | |

| fps | 71 |

| iterations | 13 |

| time\_elapsed | 372 |

| total\_timesteps | 26624 |

| train/ | |

| approx\_kl | 0.17341083 |

| clip\_fraction | 0.66 |

| clip\_range | 0.2 |

| entropy\_loss | -45.3 |

| explained\_variance | -0.76125586 |

| learning\_rate | 0.0003 |

| loss | -0.598 |

| n\_updates | 120 |

| policy\_gradient\_loss | -0.114 |

| std | 1.1 |

| value\_loss | 0.000373 |

-----------------------------------------

------------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.31 |

| time/ | |

| fps | 69 |

| iterations | 14 |

| time\_elapsed | 410 |

| total\_timesteps | 28672 |

| train/ | |

| approx\_kl | 0.15696345 |

| clip\_fraction | 0.653 |

| clip\_range | 0.2 |

| entropy\_loss | -45.5 |

| explained\_variance | -0.092641115 |

| learning\_rate | 0.0003 |

| loss | -0.592 |

| n\_updates | 130 |

| policy\_gradient\_loss | -0.113 |

| std | 1.11 |

| value\_loss | 0.000376 |

------------------------------------------

-----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.4 |

| time/ | |

| fps | 69 |

| iterations | 15 |

| time\_elapsed | 439 |

| total\_timesteps | 30720 |

| train/ | |

| approx\_kl | 0.16961072 |

| clip\_fraction | 0.649 |

| clip\_range | 0.2 |

| entropy\_loss | -45.8 |

| explained\_variance | 0.037286878 |

| learning\_rate | 0.0003 |

| loss | -0.596 |

| n\_updates | 140 |

| policy\_gradient\_loss | -0.111 |

| std | 1.12 |

| value\_loss | 0.000337 |

-----------------------------------------

-----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.48 |

| time/ | |

| fps | 69 |

| iterations | 16 |

| time\_elapsed | 474 |

| total\_timesteps | 32768 |

| train/ | |

| approx\_kl | 0.17074445 |

| clip\_fraction | 0.659 |

| clip\_range | 0.2 |

| entropy\_loss | -46.1 |

| explained\_variance | -0.74524987 |

| learning\_rate | 0.0003 |

| loss | -0.606 |

| n\_updates | 150 |

| policy\_gradient\_loss | -0.113 |

| std | 1.13 |

| value\_loss | 0.000289 |

-----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.55 |

| time/ | |

| fps | 68 |

| iterations | 17 |

| time\_elapsed | 508 |

| total\_timesteps | 34816 |

| train/ | |

| approx\_kl | 0.15214992 |

| clip\_fraction | 0.635 |

| clip\_range | 0.2 |

| entropy\_loss | -46.3 |

| explained\_variance | -0.6055962 |

| learning\_rate | 0.0003 |

| loss | -0.6 |

| n\_updates | 160 |

| policy\_gradient\_loss | -0.104 |

| std | 1.14 |

| value\_loss | 0.000344 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.63 |

| time/ | |

| fps | 67 |

| iterations | 18 |

| time\_elapsed | 547 |

| total\_timesteps | 36864 |

| train/ | |

| approx\_kl | 0.15226382 |

| clip\_fraction | 0.635 |

| clip\_range | 0.2 |

| entropy\_loss | -46.6 |

| explained\_variance | -0.3980403 |

| learning\_rate | 0.0003 |

| loss | -0.613 |

| n\_updates | 170 |

| policy\_gradient\_loss | -0.105 |

| std | 1.15 |

| value\_loss | 0.000308 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.72 |

| time/ | |

| fps | 67 |

| iterations | 19 |

| time\_elapsed | 573 |

| total\_timesteps | 38912 |

| train/ | |

| approx\_kl | 0.17435819 |

| clip\_fraction | 0.659 |

| clip\_range | 0.2 |

| entropy\_loss | -46.9 |

| explained\_variance | 0.2911234 |

| learning\_rate | 0.0003 |

| loss | -0.607 |

| n\_updates | 180 |

| policy\_gradient\_loss | -0.109 |

| std | 1.16 |

| value\_loss | 0.000241 |

----------------------------------------

--------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.79 |

| time/ | |

| fps | 66 |

| iterations | 20 |

| time\_elapsed | 619 |

| total\_timesteps | 40960 |

| train/ | |

| approx\_kl | 0.177959 |

| clip\_fraction | 0.659 |

| clip\_range | 0.2 |

| entropy\_loss | -47.1 |

| explained\_variance | 0.324762 |

| learning\_rate | 0.0003 |

| loss | -0.619 |

| n\_updates | 190 |

| policy\_gradient\_loss | -0.114 |

| std | 1.17 |

| value\_loss | 0.000239 |

--------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.91 |

| time/ | |

| fps | 65 |

| iterations | 21 |

| time\_elapsed | 656 |

| total\_timesteps | 43008 |

| train/ | |

| approx\_kl | 0.17114906 |

| clip\_fraction | 0.649 |

| clip\_range | 0.2 |

| entropy\_loss | -47.4 |

| explained\_variance | 0.43078643 |

| learning\_rate | 0.0003 |

| loss | -0.615 |

| n\_updates | 200 |

| policy\_gradient\_loss | -0.11 |

| std | 1.18 |

| value\_loss | 0.000282 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 6.99 |

| time/ | |

| fps | 64 |

| iterations | 22 |

| time\_elapsed | 693 |

| total\_timesteps | 45056 |

| train/ | |

| approx\_kl | 0.18460706 |

| clip\_fraction | 0.664 |

| clip\_range | 0.2 |

| entropy\_loss | -47.8 |

| explained\_variance | 0.32700938 |

| learning\_rate | 0.0003 |

| loss | -0.622 |

| n\_updates | 210 |

| policy\_gradient\_loss | -0.113 |

| std | 1.2 |

| value\_loss | 0.000267 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.07 |

| time/ | |

| fps | 64 |

| iterations | 23 |

| time\_elapsed | 727 |

| total\_timesteps | 47104 |

| train/ | |

| approx\_kl | 0.17804107 |

| clip\_fraction | 0.664 |

| clip\_range | 0.2 |

| entropy\_loss | -48.1 |

| explained\_variance | 0.48330164 |

| learning\_rate | 0.0003 |

| loss | -0.629 |

| n\_updates | 220 |

| policy\_gradient\_loss | -0.114 |

| std | 1.21 |

| value\_loss | 0.000258 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.15 |

| time/ | |

| fps | 63 |

| iterations | 24 |

| time\_elapsed | 768 |

| total\_timesteps | 49152 |

| train/ | |

| approx\_kl | 0.18276206 |

| clip\_fraction | 0.649 |

| clip\_range | 0.2 |

| entropy\_loss | -48.3 |

| explained\_variance | 0.39312232 |

| learning\_rate | 0.0003 |

| loss | -0.619 |

| n\_updates | 230 |

| policy\_gradient\_loss | -0.111 |

| std | 1.22 |

| value\_loss | 0.000219 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.24 |

| time/ | |

| fps | 63 |

| iterations | 25 |

| time\_elapsed | 810 |

| total\_timesteps | 51200 |

| train/ | |

| approx\_kl | 0.17275748 |

| clip\_fraction | 0.643 |

| clip\_range | 0.2 |

| entropy\_loss | -48.6 |

| explained\_variance | 0.30047262 |

| learning\_rate | 0.0003 |

| loss | -0.617 |

| n\_updates | 240 |

| policy\_gradient\_loss | -0.109 |

| std | 1.23 |

| value\_loss | 0.000211 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.3 |

| time/ | |

| fps | 63 |

| iterations | 26 |

| time\_elapsed | 843 |

| total\_timesteps | 53248 |

| train/ | |

| approx\_kl | 0.16315487 |

| clip\_fraction | 0.66 |

| clip\_range | 0.2 |

| entropy\_loss | -48.8 |

| explained\_variance | 0.49055022 |

| learning\_rate | 0.0003 |

| loss | -0.624 |

| n\_updates | 250 |

| policy\_gradient\_loss | -0.112 |

| std | 1.24 |

| value\_loss | 0.000197 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.38 |

| time/ | |

| fps | 63 |

| iterations | 27 |

| time\_elapsed | 870 |

| total\_timesteps | 55296 |

| train/ | |

| approx\_kl | 0.16986074 |

| clip\_fraction | 0.656 |

| clip\_range | 0.2 |

| entropy\_loss | -49.1 |

| explained\_variance | 0.2295754 |

| learning\_rate | 0.0003 |

| loss | -0.621 |

| n\_updates | 260 |

| policy\_gradient\_loss | -0.11 |

| std | 1.25 |

| value\_loss | 0.000167 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.46 |

| time/ | |

| fps | 63 |

| iterations | 28 |

| time\_elapsed | 904 |

| total\_timesteps | 57344 |

| train/ | |

| approx\_kl | 0.16871276 |

| clip\_fraction | 0.652 |

| clip\_range | 0.2 |

| entropy\_loss | -49.4 |

| explained\_variance | 0.43809032 |

| learning\_rate | 0.0003 |

| loss | -0.623 |

| n\_updates | 270 |

| policy\_gradient\_loss | -0.108 |

| std | 1.26 |

| value\_loss | 0.000191 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.54 |

| time/ | |

| fps | 62 |

| iterations | 29 |

| time\_elapsed | 946 |

| total\_timesteps | 59392 |

| train/ | |

| approx\_kl | 0.17472066 |

| clip\_fraction | 0.651 |

| clip\_range | 0.2 |

| entropy\_loss | -49.6 |

| explained\_variance | 0.53492916 |

| learning\_rate | 0.0003 |

| loss | -0.64 |

| n\_updates | 280 |

| policy\_gradient\_loss | -0.11 |

| std | 1.27 |

| value\_loss | 0.000207 |

----------------------------------------

---------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.63 |

| time/ | |

| fps | 62 |

| iterations | 30 |

| time\_elapsed | 986 |

| total\_timesteps | 61440 |

| train/ | |

| approx\_kl | 0.2177889 |

| clip\_fraction | 0.668 |

| clip\_range | 0.2 |

| entropy\_loss | -49.8 |

| explained\_variance | 0.4475624 |

| learning\_rate | 0.0003 |

| loss | -0.647 |

| n\_updates | 290 |

| policy\_gradient\_loss | -0.113 |

| std | 1.28 |

| value\_loss | 0.000249 |

---------------------------------------

---------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.71 |

| time/ | |

| fps | 62 |

| iterations | 31 |

| time\_elapsed | 1018 |

| total\_timesteps | 63488 |

| train/ | |

| approx\_kl | 0.2127443 |

| clip\_fraction | 0.683 |

| clip\_range | 0.2 |

| entropy\_loss | -50.1 |

| explained\_variance | 0.5402695 |

| learning\_rate | 0.0003 |

| loss | -0.636 |

| n\_updates | 300 |

| policy\_gradient\_loss | -0.109 |

| std | 1.3 |

| value\_loss | 0.000288 |

---------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.79 |

| time/ | |

| fps | 62 |

| iterations | 32 |

| time\_elapsed | 1050 |

| total\_timesteps | 65536 |

| train/ | |

| approx\_kl | 0.20365806 |

| clip\_fraction | 0.668 |

| clip\_range | 0.2 |

| entropy\_loss | -50.5 |

| explained\_variance | 0.38361663 |

| learning\_rate | 0.0003 |

| loss | -0.654 |

| n\_updates | 310 |

| policy\_gradient\_loss | -0.109 |

| std | 1.31 |

| value\_loss | 0.000288 |

----------------------------------------

---------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.88 |

| time/ | |

| fps | 62 |

| iterations | 33 |

| time\_elapsed | 1086 |

| total\_timesteps | 67584 |

| train/ | |

| approx\_kl | 0.1695806 |

| clip\_fraction | 0.652 |

| clip\_range | 0.2 |

| entropy\_loss | -50.7 |

| explained\_variance | 0.5750711 |

| learning\_rate | 0.0003 |

| loss | -0.657 |

| n\_updates | 320 |

| policy\_gradient\_loss | -0.108 |

| std | 1.32 |

| value\_loss | 0.000243 |

---------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 7.99 |

| time/ | |

| fps | 61 |

| iterations | 34 |

| time\_elapsed | 1126 |

| total\_timesteps | 69632 |

| train/ | |

| approx\_kl | 0.19588318 |

| clip\_fraction | 0.68 |

| clip\_range | 0.2 |

| entropy\_loss | -51 |

| explained\_variance | 0.5569705 |

| learning\_rate | 0.0003 |

| loss | -0.65 |

| n\_updates | 330 |

| policy\_gradient\_loss | -0.115 |

| std | 1.33 |

| value\_loss | 0.000501 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.07 |

| time/ | |

| fps | 61 |

| iterations | 35 |

| time\_elapsed | 1171 |

| total\_timesteps | 71680 |

| train/ | |

| approx\_kl | 0.20773962 |

| clip\_fraction | 0.681 |

| clip\_range | 0.2 |

| entropy\_loss | -51.2 |

| explained\_variance | 0.43216604 |

| learning\_rate | 0.0003 |

| loss | -0.651 |

| n\_updates | 340 |

| policy\_gradient\_loss | -0.114 |

| std | 1.34 |

| value\_loss | 0.000387 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.15 |

| time/ | |

| fps | 60 |

| iterations | 36 |

| time\_elapsed | 1209 |

| total\_timesteps | 73728 |

| train/ | |

| approx\_kl | 0.27089703 |

| clip\_fraction | 0.703 |

| clip\_range | 0.2 |

| entropy\_loss | -51.5 |

| explained\_variance | 0.4767074 |

| learning\_rate | 0.0003 |

| loss | -0.655 |

| n\_updates | 350 |

| policy\_gradient\_loss | -0.116 |

| std | 1.36 |

| value\_loss | 0.000456 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.22 |

| time/ | |

| fps | 60 |

| iterations | 37 |

| time\_elapsed | 1251 |

| total\_timesteps | 75776 |

| train/ | |

| approx\_kl | 0.20499143 |

| clip\_fraction | 0.68 |

| clip\_range | 0.2 |

| entropy\_loss | -51.8 |

| explained\_variance | 0.49647385 |

| learning\_rate | 0.0003 |

| loss | -0.673 |

| n\_updates | 360 |

| policy\_gradient\_loss | -0.109 |

| std | 1.37 |

| value\_loss | 0.000285 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.31 |

| time/ | |

| fps | 60 |

| iterations | 38 |

| time\_elapsed | 1294 |

| total\_timesteps | 77824 |

| train/ | |

| approx\_kl | 0.21018618 |

| clip\_fraction | 0.667 |

| clip\_range | 0.2 |

| entropy\_loss | -52.1 |

| explained\_variance | 0.5987345 |

| learning\_rate | 0.0003 |

| loss | -0.668 |

| n\_updates | 370 |

| policy\_gradient\_loss | -0.109 |

| std | 1.38 |

| value\_loss | 0.000383 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.38 |

| time/ | |

| fps | 59 |

| iterations | 39 |

| time\_elapsed | 1335 |

| total\_timesteps | 79872 |

| train/ | |

| approx\_kl | 0.199864 |

| clip\_fraction | 0.664 |

| clip\_range | 0.2 |

| entropy\_loss | -52.4 |

| explained\_variance | 0.73710966 |

| learning\_rate | 0.0003 |

| loss | -0.641 |

| n\_updates | 380 |

| policy\_gradient\_loss | -0.108 |

| std | 1.4 |

| value\_loss | 0.000469 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.46 |

| time/ | |

| fps | 59 |

| iterations | 40 |

| time\_elapsed | 1371 |

| total\_timesteps | 81920 |

| train/ | |

| approx\_kl | 0.21693608 |

| clip\_fraction | 0.67 |

| clip\_range | 0.2 |

| entropy\_loss | -52.6 |

| explained\_variance | 0.6515416 |

| learning\_rate | 0.0003 |

| loss | -0.664 |

| n\_updates | 390 |

| policy\_gradient\_loss | -0.113 |

| std | 1.4 |

| value\_loss | 0.000322 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.56 |

| time/ | |

| fps | 59 |

| iterations | 41 |

| time\_elapsed | 1414 |

| total\_timesteps | 83968 |

| train/ | |

| approx\_kl | 0.25370413 |

| clip\_fraction | 0.703 |

| clip\_range | 0.2 |

| entropy\_loss | -52.9 |

| explained\_variance | 0.61931443 |

| learning\_rate | 0.0003 |

| loss | -0.682 |

| n\_updates | 400 |

| policy\_gradient\_loss | -0.117 |

| std | 1.42 |

| value\_loss | 0.000613 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.68 |

| time/ | |

| fps | 58 |

| iterations | 42 |

| time\_elapsed | 1467 |

| total\_timesteps | 86016 |

| train/ | |

| approx\_kl | 0.20837772 |

| clip\_fraction | 0.68 |

| clip\_range | 0.2 |

| entropy\_loss | -53.2 |

| explained\_variance | 0.57892764 |

| learning\_rate | 0.0003 |

| loss | -0.681 |

| n\_updates | 410 |

| policy\_gradient\_loss | -0.109 |

| std | 1.43 |

| value\_loss | 0.000396 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.75 |

| time/ | |

| fps | 58 |

| iterations | 43 |

| time\_elapsed | 1499 |

| total\_timesteps | 88064 |

| train/ | |

| approx\_kl | 0.21504544 |

| clip\_fraction | 0.69 |

| clip\_range | 0.2 |

| entropy\_loss | -53.5 |

| explained\_variance | 0.6609482 |

| learning\_rate | 0.0003 |

| loss | -0.679 |

| n\_updates | 420 |

| policy\_gradient\_loss | -0.113 |

| std | 1.45 |

| value\_loss | 0.000463 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.83 |

| time/ | |

| fps | 58 |

| iterations | 44 |

| time\_elapsed | 1547 |

| total\_timesteps | 90112 |

| train/ | |

| approx\_kl | 0.22417668 |

| clip\_fraction | 0.677 |

| clip\_range | 0.2 |

| entropy\_loss | -53.7 |

| explained\_variance | 0.6785698 |

| learning\_rate | 0.0003 |

| loss | -0.679 |

| n\_updates | 430 |

| policy\_gradient\_loss | -0.112 |

| std | 1.46 |

| value\_loss | 0.000392 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.89 |

| time/ | |

| fps | 57 |

| iterations | 45 |

| time\_elapsed | 1592 |

| total\_timesteps | 92160 |

| train/ | |

| approx\_kl | 0.30695316 |

| clip\_fraction | 0.727 |

| clip\_range | 0.2 |

| entropy\_loss | -54 |

| explained\_variance | 0.3916331 |

| learning\_rate | 0.0003 |

| loss | -0.686 |

| n\_updates | 440 |

| policy\_gradient\_loss | -0.118 |

| std | 1.48 |

| value\_loss | 0.000427 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 8.96 |

| time/ | |

| fps | 57 |

| iterations | 46 |

| time\_elapsed | 1634 |

| total\_timesteps | 94208 |

| train/ | |

| approx\_kl | 0.2143324 |

| clip\_fraction | 0.675 |

| clip\_range | 0.2 |

| entropy\_loss | -54.4 |

| explained\_variance | 0.61613464 |

| learning\_rate | 0.0003 |

| loss | -0.671 |

| n\_updates | 450 |

| policy\_gradient\_loss | -0.112 |

| std | 1.49 |

| value\_loss | 0.000318 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 9.02 |

| time/ | |

| fps | 57 |

| iterations | 47 |

| time\_elapsed | 1677 |

| total\_timesteps | 96256 |

| train/ | |

| approx\_kl | 0.21197191 |

| clip\_fraction | 0.697 |

| clip\_range | 0.2 |

| entropy\_loss | -54.7 |

| explained\_variance | 0.63544333 |

| learning\_rate | 0.0003 |

| loss | -0.691 |

| n\_updates | 460 |

| policy\_gradient\_loss | -0.117 |

| std | 1.51 |

| value\_loss | 0.000437 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 9.08 |

| time/ | |

| fps | 57 |

| iterations | 48 |

| time\_elapsed | 1715 |

| total\_timesteps | 98304 |

| train/ | |

| approx\_kl | 0.20541595 |

| clip\_fraction | 0.678 |

| clip\_range | 0.2 |

| entropy\_loss | -54.9 |

| explained\_variance | 0.5891249 |

| learning\_rate | 0.0003 |

| loss | -0.696 |

| n\_updates | 470 |

| policy\_gradient\_loss | -0.113 |

| std | 1.52 |

| value\_loss | 0.000378 |

----------------------------------------

----------------------------------------

| rollout/ | |

| ep\_len\_mean | 1e+03 |

| ep\_rew\_mean | 9.14 |

| time/ | |

| fps | 56 |

| iterations | 49 |

| time\_elapsed | 1766 |

| total\_timesteps | 100352 |

| train/ | |

| approx\_kl | 0.21476486 |

| clip\_fraction | 0.664 |

| clip\_range | 0.2 |

| entropy\_loss | -55.2 |

| explained\_variance | 0.6742269 |

| learning\_rate | 0.0003 |

| loss | -0.702 |

| n\_updates | 480 |

| policy\_gradient\_loss | -0.111 |

| std | 1.53 |

| value\_loss | 0.000297 |

----------------------------------------

训练完成！模型已保存至: selftouch\_ppo\_model.zip

开始评估模型...

Episode 1: 总奖励=18.12

Episode 2: 总奖励=18.15

Episode 3: 总奖励=19.93

Episode 4: 总奖励=17.21

Episode 5: 总奖励=18.66

Episode 6: 总奖励=21.59

Episode 7: 总奖励=19.23

Episode 8: 总奖励=25.37

Episode 9: 总奖励=18.16

Episode 10: 总奖励=19.60

平均奖励: 19.60 ± 2.25