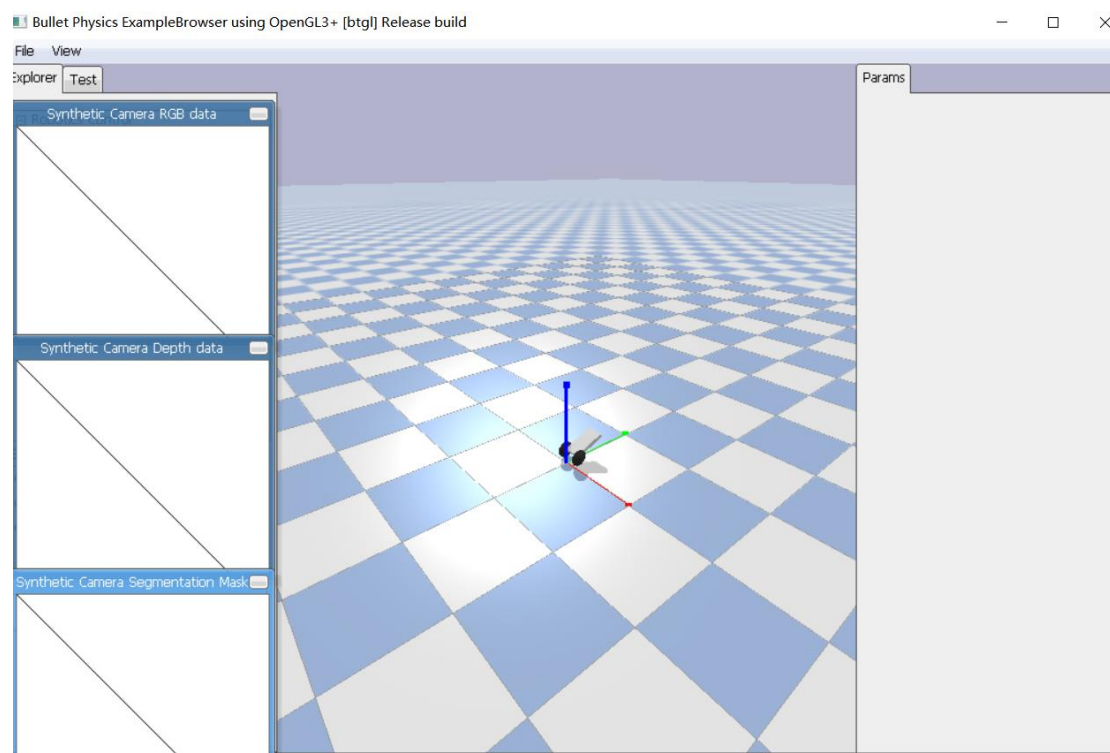


Parameters:

```
env,  
network='mlp',  
hiddens=[15,11],  
lr=1.5e-3,  
total_timesteps=100000,  
buffer_size=100000,  
exploration_fraction=0.2,  
exploration_final_eps=0.03,  
print_freq=10,
```

The bot is move with random speed and may fall due to weight.



Some screenshots when running (IDE:PYCHARM)

Logging to C:\Users\P65\AppData\Local\Temp\openai-2021-08-

```
-----  
| % time spent exploring | 98      |  
| episodes               | 10      |  
| mean 100 episode reward | 4       |  
| steps                  | 410     |  
-----
```

```
-----  
| % time spent exploring | 95      |  
| episodes               | 20      |  
| mean 100 episode reward | 3.8     |  
| steps                  | 832     |  
-----
```

```
-----  
| % time spent exploring | 94      |  
| episodes               | 30      |  
| mean 100 episode reward | 3.4     |  
| steps                  | 1.17e+03 |  
-----
```

```
-----  
| % time spent exploring | 92      |  
| episodes               | 40      |  
| mean 100 episode reward | 3.4     |  
| steps                  | 1.56e+03 |  
-----
```

```
| steps | 8.51e+04 |
```

```
-----  
| % time spent exploring | 3 |  
| episodes | 400 |  
| mean 100 episode reward | 57.6 |  
| steps | 8.92e+04 |  
-----
```

Saving model due to mean reward increase: 50.9 -> 57.5

```
-----  
| % time spent exploring | 3 |  
| episodes | 410 |  
| mean 100 episode reward | 60.5 |  
| steps | 9.59e+04 |  
-----
```

Restored model with mean reward: 57.5

numActiveThreads = 0

stopping threads

Thread with taskId 0 with handle 00000000000001020 exiting

Thread TERMINATED

finished

numActiveThreads = 0

btShutDownExampleBrowser stopping threads

Thread with taskId 0 with handle 0000000000000C98 exiting

Thread TERMINATED

episodes	110	
mean 100 episode reward	3.4	
steps	4.36e+03	

% time spent exploring	77	
episodes	120	
mean 100 episode reward	3.4	
steps	4.74e+03	

% time spent exploring	75	
episodes	130	
mean 100 episode reward	3.4	
steps	5.06e+03	

% time spent exploring	73	
episodes	140	
mean 100 episode reward	3.4	
steps	5.47e+03	

% time spent exploring	71	
episodes	150	
mean 100 episode reward	3.4	
steps	5.81e+03	

Saving model due to mean reward increase: 31.8 -> 39.9

Saving model due to mean reward increase: 39.9 -> 46.2

% time spent exploring	2	
episodes	280	
mean 100 episode reward	50.6	
steps	8.5e+04	

Saving model due to mean reward increase: 46.2 -> 54.4

% time spent exploring	2	
episodes	290	
mean 100 episode reward	60.1	
steps	9.78e+04	

Restored model with mean reward: 54.4

Saving model to balance.pkl

numActiveThreads = 0

stopping threads

Thread with taskId 0 with handle 00000000000001040 exiting

Thread TERMINATED

finished

numActiveThreads = 0

btShutDownExampleBrowser stopping threads

Thread with taskId 0 with handle 0000000000000174 exiting

Thread TERMINATED