



强化学习机械臂控制大作业说明

自动化与感知学院

2025年10月



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



智能机器人与机器视觉实验室
Intelligent Robotics and Machine Vision Laboratory

irmv.sjtu.edu.cn

- 大作业目标
- 任务介绍
 - 任务目标
 - 任务环境
- 仿真环境搭建
 - Pybullet仿真引擎
 - Gym标准强化学习仿真环境
- 抓取任务训练
 - Pytorch深度学习框架
 - 网络结构
 - 服务器的使用
 - 大作业：仿真抓取

实践目标

利用仿真环境（PyBullet）设置基于Gym的强化学习仿真环境，
学习强化学习仿真环境的重要组成部分（观察、动作、奖励、环境交互），
训练强化学习网络，完成机械臂抓取任务的路径规划。

能力提升

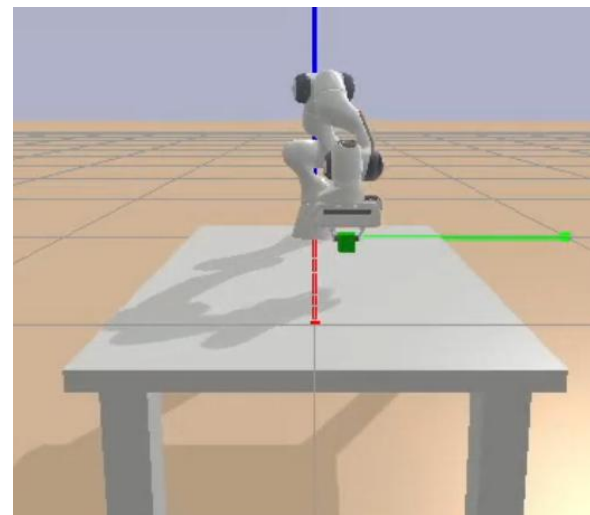
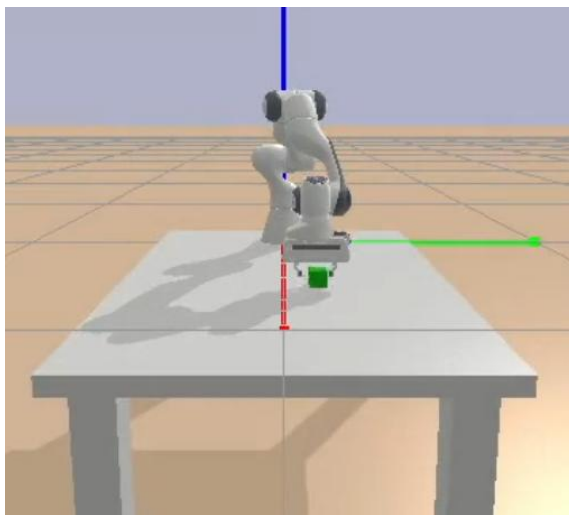
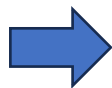
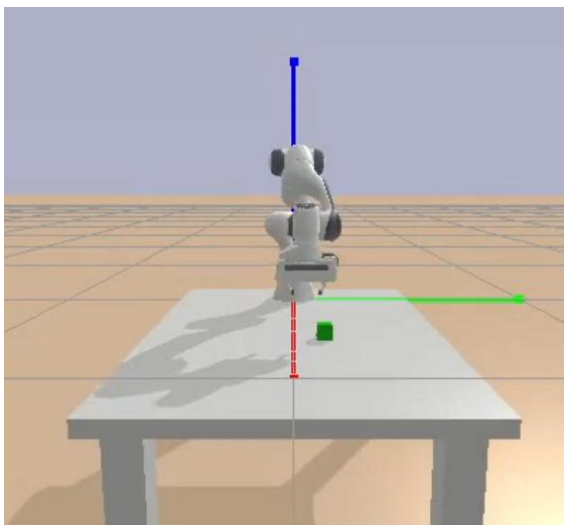
掌握强化学习的基本概念；
熟悉强化学习在机器人控制领域的应用；
提高编程实践能力和问题解决能力。

任务目标

任务目标

使用仿真环境pybullet完善抓取方块任务的仿真环境，并使用强化学习算法（DQN），完成抓取方块任务的操作策略的训练和优化。

为简化任务设置，减少训练时间，本次大作业只需要完成靠近目标物体的任务。即机械臂夹爪距离目标物体的距离小于阈值（0.005m）则视为任务成功。



任务环境

智能体：

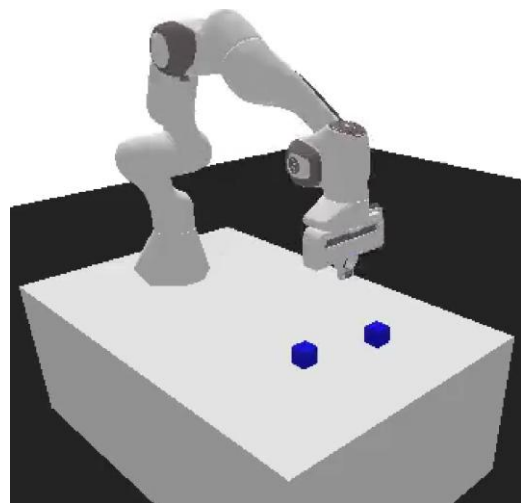
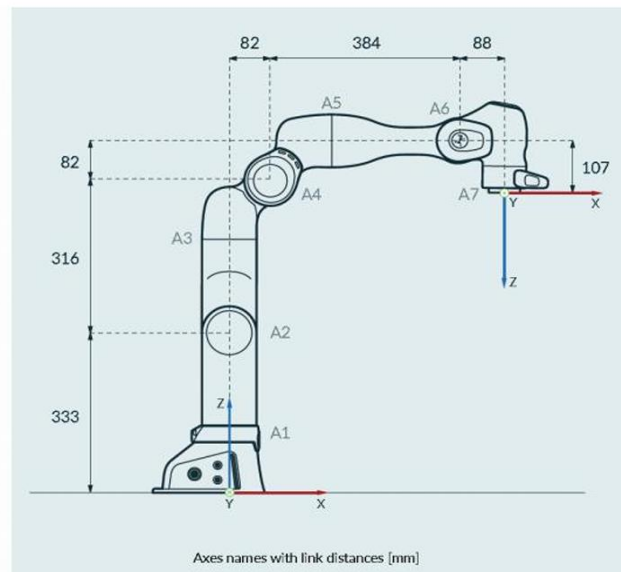
Franka Emika Panda机器人，7-轴机械臂，它的规格为3kg载重，850mm臂展。

观察：

- 机械臂关节位置（9维）：7维机械臂关节+2维夹爪
- 机械臂夹爪位姿：7维（3维位置+4维四元数）
- 机械臂夹爪与目标物体的距离之差：3维
- 目标物体位姿：7维（3维位置+4维四元数）

动作（离散空间）：

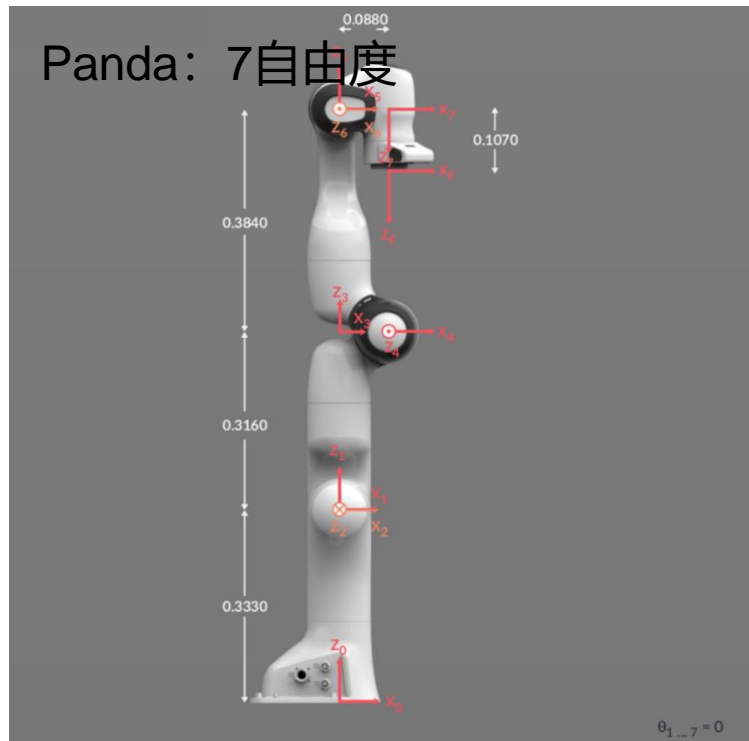
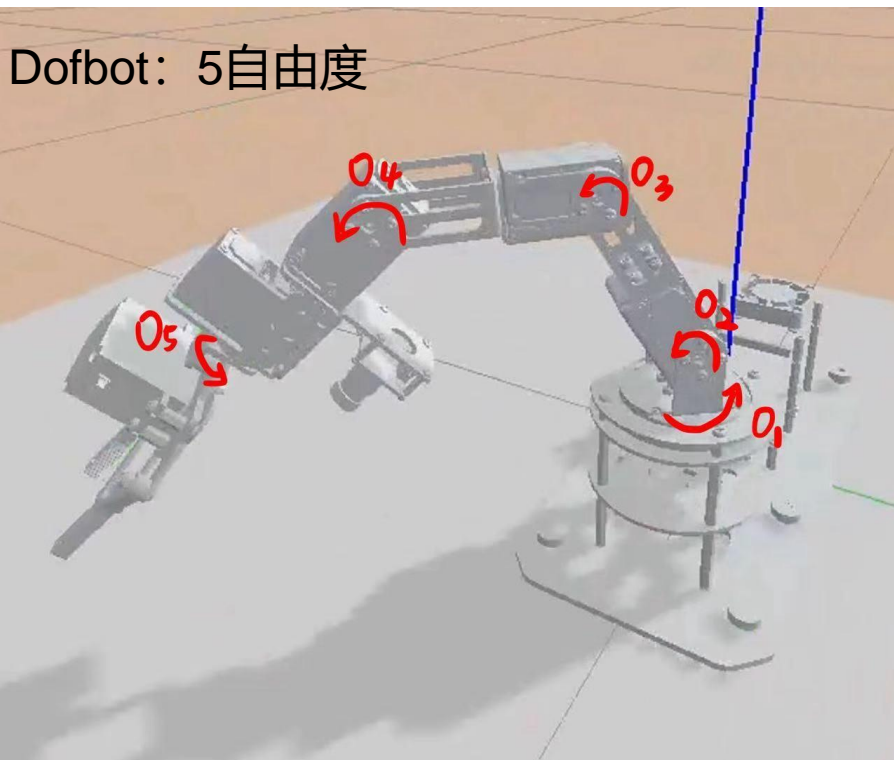
- 动作1：夹爪沿x轴负方向移动0.01m
- 动作2：夹爪沿x轴正方向移动0.01m
- 动作3：夹爪沿y轴负方向移动0.01m
- 动作4：夹爪沿y轴正方向移动0.01m
- 动作5：夹爪沿z轴负方向移动0.01m
- 动作6：夹爪沿z轴正方向移动0.01m
- 动作7：保持静止



机械臂自由度

动作空间：笛卡尔空间（末端夹爪的位姿）

控制逻辑：输入动作→末端位姿→（逆运动学）关节位置

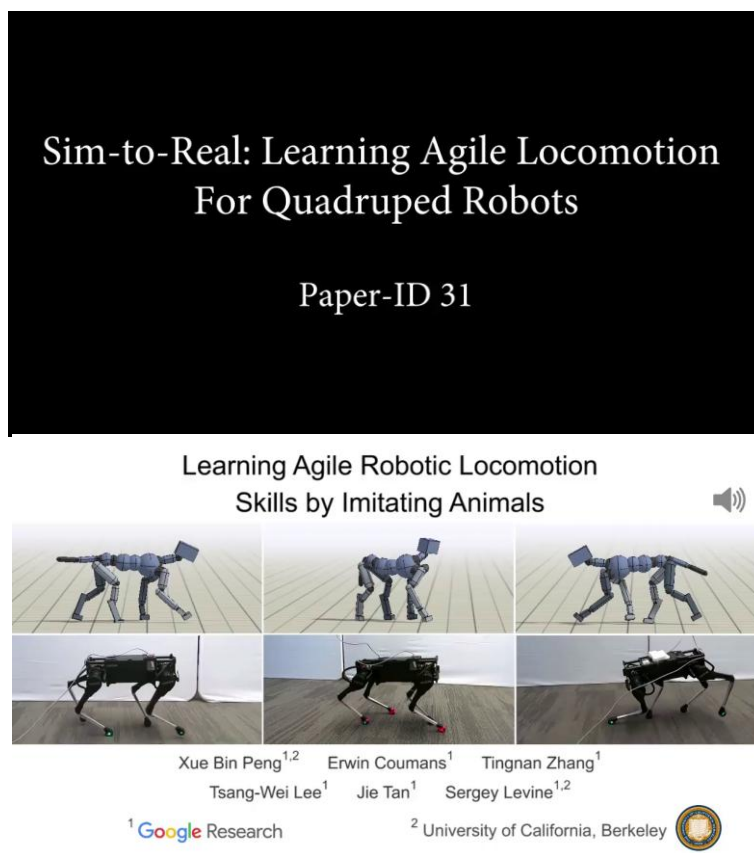
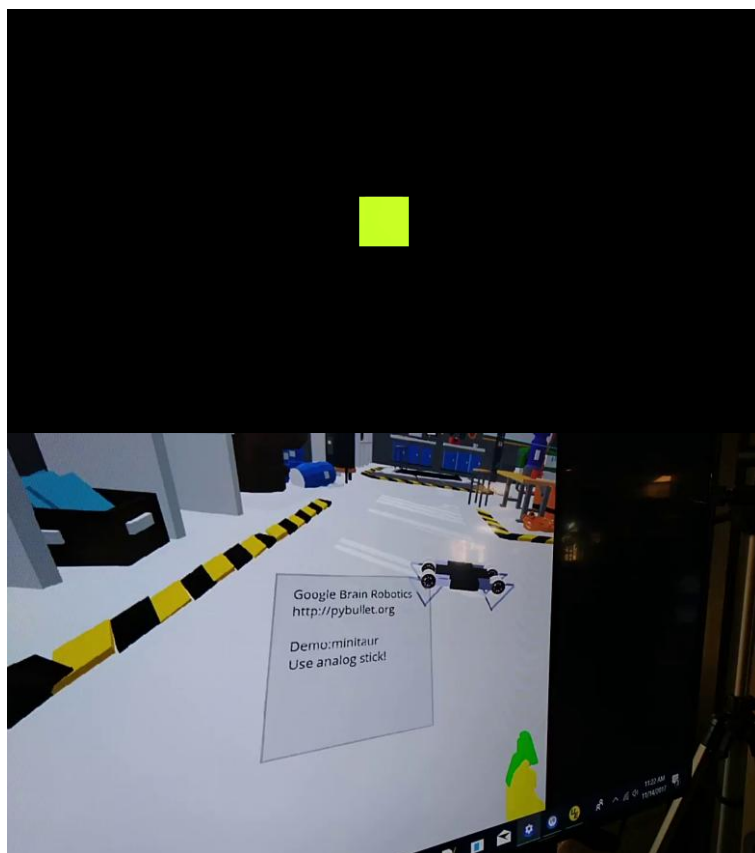


RL动作采样逻辑：在笛卡尔空间中随机采样

自由度不足：逆运动学可能无解

PyBullet 基于著名的开源物理引擎 bullet 开发，封装成了 Python 的一个模块，用于机器人仿真和学习。PyBullet 支持加载 URDF、SDF、MJCF 等多种机器人描述文件，并提供正/逆向运动学、正/逆向动力学、碰撞检测等功能。(https://pybullet.org/wordpress/)，Bullet 物理 SDK 包括 PyBullet 机器人示例，如模拟的 Minitaur 四足机器人、使用 TensorFlow 推理的仿人机器人跑步和 KUKA 机械臂抓取物体。

PyBullet安装: `pip install pybullet`



Gym环境

Gym介绍

Gym 是 OpenAI 提供的一个用于开发和测试强化学习算法的工具库，包含多种标准化的环境接口（如机器人控制、游戏等）。它通过统一的 API，方便用户创建、交互和评估各种强化学习任务。

Gym环境组成

1. 状态空间 (observation space): 描述环境的状态，例如位置、速度等。状态可以是连续的或离散的。
2. 动作空间 (action space): 描述智能体可以采取的动作，例如移动方向、速度等。动作空间也可以是离散或连续的。
3. 环境接口
 - `env.reset()`: 重置环境，返回初始状态。
 - `env.step(action)`: 执行动作，返回下一个状态、奖励、是否结束和额外信息。
 - `env.render()`: 渲染环境，用于可视化。
 - `env.close()`: 关闭环境，释放资源。

深度学习框架Pytorch



PyTorch 是一个开源的机器学习库，主要用于进行计算机视觉（CV）、自然语言处理（NLP）、语音识别等领域的研究和开发。

◆ 张量操作：

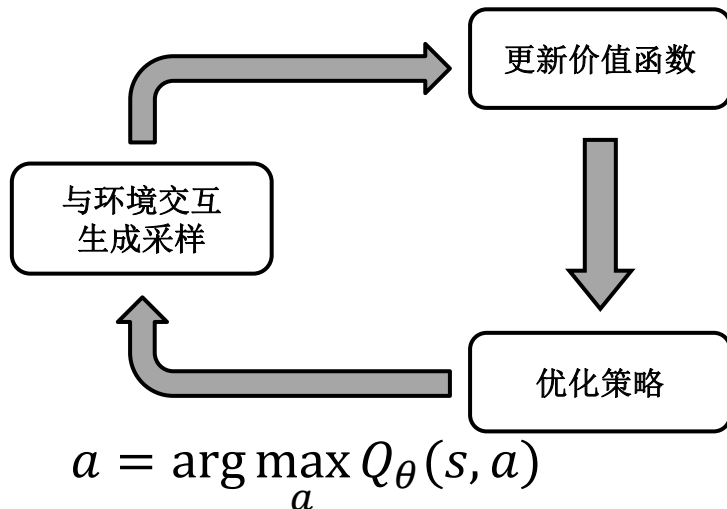
- 创建张量：`torch.tensor(data)`；`torch.rand(size)`
- 张量属性：张量形状`.shape`；张量的数据类型`.dtype`；张量所在设备`.device`
- 形状操作：矩阵乘法`torch.matmul(x, y)`；返回最大值的索引`torch.argmax(x, dim)`；计算softmax `torch.softmax(x, dim)`

◆ torch.nn 模块：构建和训练神经网络的核心模块

- `nn.Module`：所有自定义神经网络模型的基类
- 损失函数：均方误差损失（`nn.MSELoss`）、交叉熵损失（`nn.CrossEntropyLoss`）等
- 容器类：`nn.Sequential`：允许将多个层按顺序组合起来，形成简单的线性堆叠网络。
- 线性层函数：`torch.nn.Linear(in_features, out_features)`
- 激活函数：`torch.nn.ReLU()`；`torch.nn.Tanh()`...

网络结构

$$Q_{\theta}(s, a) \leftarrow r(s, a) + \gamma \max_{a'} Q_{\theta}(s', a')$$



更新价值函数:

```
# update target network
if global_step % args.target_network_frequency == 0:
    for target_network_param, q_network_param in zip(target_network.parameters(), q_network.parameters()):
        target_network_param.data.copy_(
            args.tau * q_network_param.data + (1.0 - args.tau) * target_network_param.data
        )
```

Q-network:

```
# ALGO LOGIC: initialize agent here:
class QNetwork(nn.Module):
    def __init__(self, env):
        super().__init__()
        self.network = nn.Sequential(
            nn.Linear(np.array(env.single_observation_space.shape).prod(), 120),
            nn.ReLU(),
            nn.Linear(120, 84),
            nn.ReLU(),
            nn.Linear(84, env.single_action_space.n),
        )

    def forward(self, x):
        return self.network(x)
```

优化策略:

```
q_values = q_network(torch.Tensor(obs).to(device))
actions = torch.argmax(q_values, dim=1).cpu().numpy()
```



作业（机械臂强化学习控制部分）



任务目标

使用仿真环境pybullet搭建抓取方块任务的仿真环境，并使用强化学习算法，完成抓取方块任务的操作策略的训练和优化。

报告要求

- 任务一：完善仿真环境中的**奖励函数**。
- 任务二：定义**动作空间和观测空间**。
- 任务三：在任务二的基础上完善**step函数**，根据输入动作下发机械臂控制指令。利用test_gym.py验证动作的可行性。
- 任务四：基于DQN算法，完成**操作策略的训练**，记录训练曲线，并测试效果。



作业（机械臂强化学习控制部分）



任务一：在代码panda_env.py中，完善仿真环境中的**奖励函数**。

提示：可以考虑以下几个方向

- **靠近奖励**：目标方块和夹爪（TCP）的距离越近，奖励越大。
- **到达奖励**：当目标方块和夹爪的距离小于目标阈值时，给予奖励。

```
# TODO: 完善reward function
def _get_reward(self):
    obs = self._get_obs_dict()
    info = self._get_info()

    reward = 0
    return reward
```

[要求]

设计合理的奖励函数，使得强化学习网络能够根据该奖励函数学习到正确的动作



作业（机械臂强化学习控制部分）



- 任务二：定义动作空间和观测空间。

[要求1] 完成环境初始化中的observation space和action space的定义。

```
# TODO: observation space
# if obs_mode == "state": ## 训练模式下使用
#     self.observation_space =
# elif obs_mode == "state_dict": ## 字典形式，方便读取数据
#     self.observation_space =

# TODO: action space
# self.action_space =
```

[要求2]完成_get_obs函数，从环境中得到观测信息。

```
def _get_obs_dict(self):
    Observation = self._panda.getObservation()

    # TODO: add suitable observations here

    return Observation
```

获取物体位姿函数：

- getBasePositionAndOrientation(objectUnique Id)：返回位置列表（包含3个浮点数）以及方向列表（包含4个浮点数，按 [x, y, z, w] 顺序排列）。



作业（机械臂强化学习控制部分）



- 任务三：在任务二的基础上完善**step函数**，根据输入动作下发机械臂控制指令。利用test_gym.py验证动作的可行性。

```
## discrete action: -dx, dx, -dy, dy, -dz, dz, static
def step(self, action):
    |
    |
    # TODO: Define suitable realAction here
    # self.realAction = np.array([dx, dy, dz, 0.04])

    if self.terminated:
        self.realAction = np.array([0, 0, 0, 0])
    self._panda.applyAction(self.realAction)
    p.stepSimulation()
    if self.render_mode == "human":
        time.sleep(self._timeStep)

    terminated = self._termination() ## task success check
    truncated = False ## step limitation
    self._observation = self._get_obs()
    reward = self._get_reward()
    info = self._get_info()

    return self._observation, reward, terminated, truncated, info
```

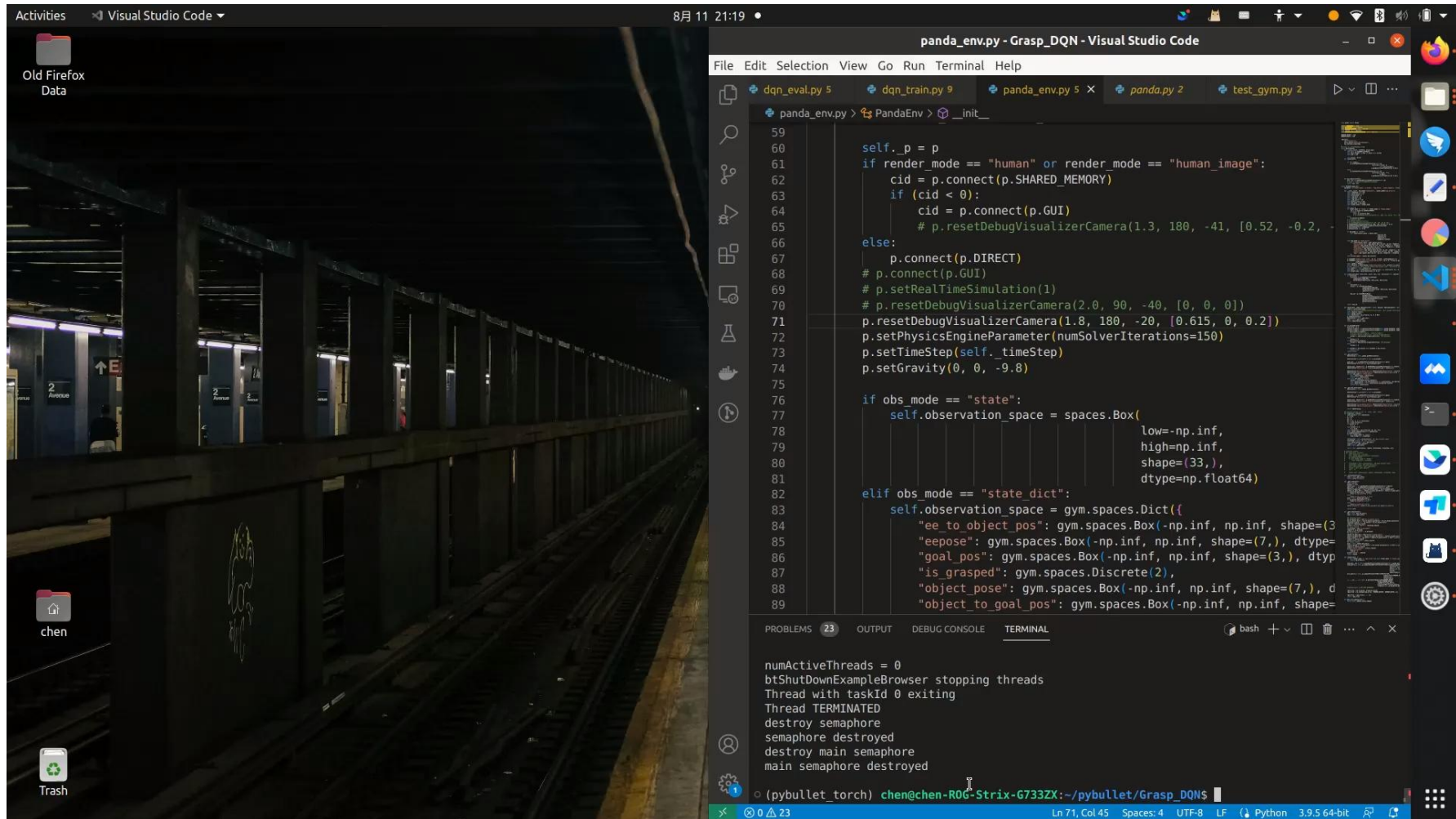
Step函数详解：

1. 将输入的离散动作转换为机械臂任务空间的动作：机械臂夹爪目标位置、方向和夹爪开合程度
2. 计算逆运动学，将机械臂夹爪位置转换为关节空间位置
3. 利用PD控制下发控制指令
4. 返回下一时间步的观测信息、任务是否完成标志以及获得的奖励。

注意：每一步位置的变化量的大小需要仔细考虑



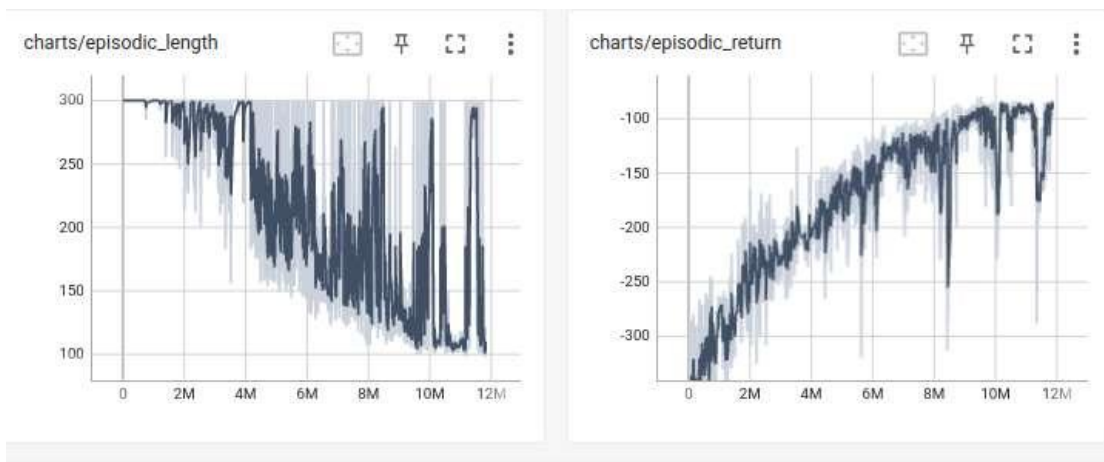
作业 (机械臂强化学习控制部分)



运行test_gym.py, 在命令行中输入动作, 在可视化窗口中验证动作设计的合理性

任务四：基于DQN算法，完成**操作策略的训练**，记录训练曲线，并测试效果。

训练曲线示例：



[要求]

1. 基于提供的DQN代码，训练抓取任务的强化学习策略
2. 使用tensorboard添加训练信息记录，并可视化训练数据
3. 测试策略效果，在提交附件中包含视频形式的结果展示

训练曲线记录工具：tensorboard

```
pip install tensorboard
```

记录事件：

```
from torch.utils.tensorboard import SummaryWriter

# 初始化SummaryWriter
writer = SummaryWriter('runs/experiment_name')
```

记录信息：

```
writer.add_scalar("losses/td_loss", loss, global_step)
writer.add_scalar("losses/q_values", old_val.mean().item(), global_step)
```

信息可视化：

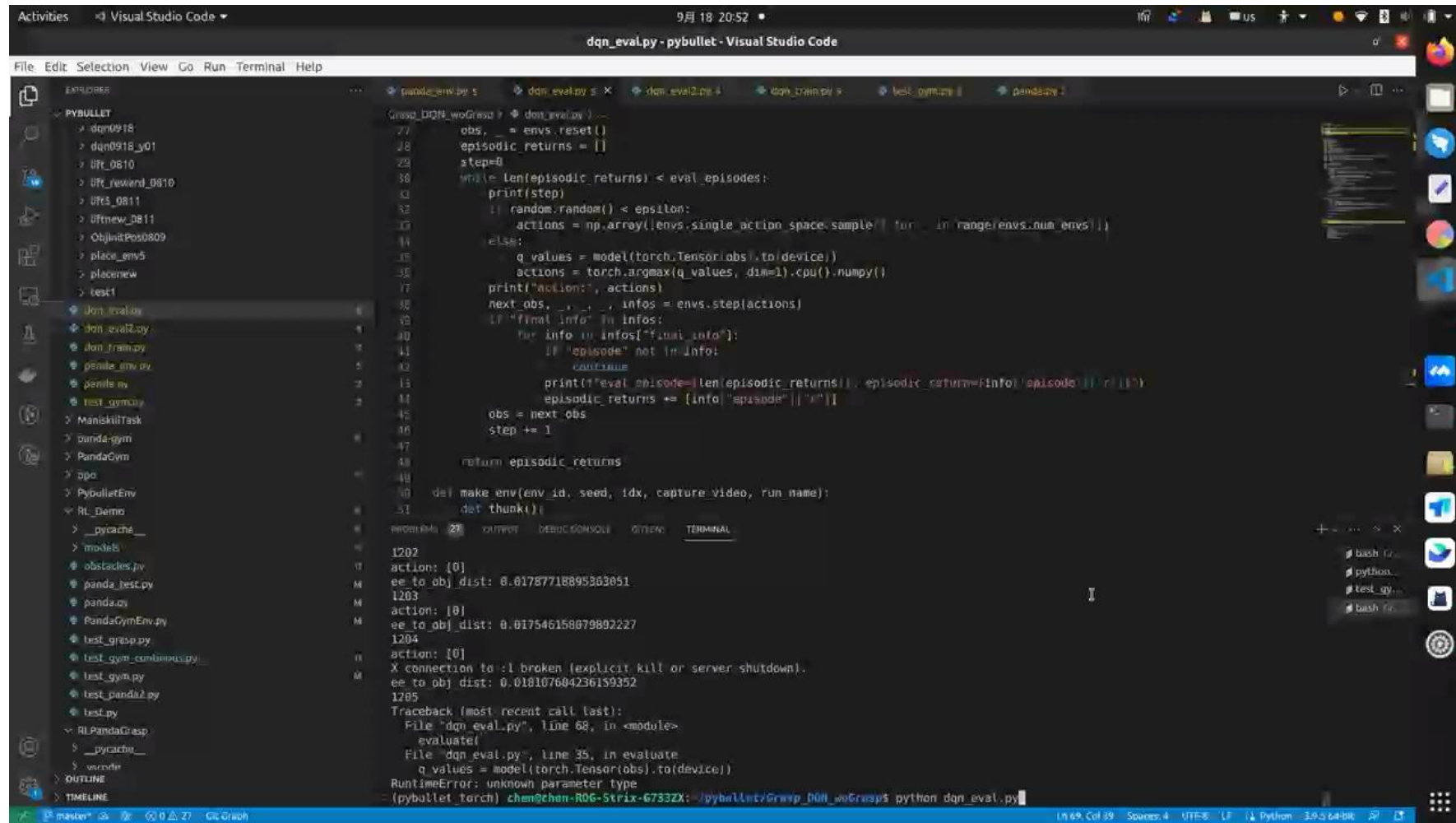
tensorboard --logdir=runs/



作业 (机械臂强化学习控制部分)



抓取策略展示



```
File Edit Selection View Go Run Terminal Help
dqn_eval.py - pybullet - Visual Studio Code
9月18 20:52
File Edit Selection View Go Run Terminal Help
EXPLORER
PYBULLET
  dqn0918
  dqn0918_y01
  Uff_0810
  Uff_reward_0810
  Uff3_0811
  Uffnew_0811
  ObjInitPos0809
  place_env5
  placenew
  test1
  dqn_eval.py
  dqn_eval2.py
  dqn_train.py
  panda_env.py
  panda_mv
  test_gym.py
  ManiskillTask
  panda-gym
  PandaGym
  dpo
  PybulletEnv
  RL_Demo
  __pycache__
  models
  obstacles.py
  panda_test.py
  panda.py
  PandaGymEnv.py
  test_grasp.py
  test_gym_combine.py
  test_gym.py
  test_panda2.py
  test.py
  RL_PandaGrasp
  __pycache__
  vncvnc
  OUTLINE
  TIMELINE
  master*
  0.27
  Gx Graph
  dqn_eval.py
  77
  78
  79
  80
  81
  82
  83
  84
  85
  86
  87
  88
  89
  90
  91
  92
  93
  94
  95
  96
  97
  98
  99
  100
  101
  102
  103
  104
  105
  106
  107
  108
  109
  110
  111
  112
  113
  114
  115
  116
  117
  118
  119
  120
  121
  122
  123
  124
  125
  126
  127
  128
  129
  130
  131
  132
  133
  134
  135
  136
  137
  138
  139
  140
  141
  142
  143
  144
  145
  146
  147
  148
  149
  150
  151
  152
  153
  154
  155
  156
  157
  158
  159
  160
  161
  162
  163
  164
  165
  166
  167
  168
  169
  170
  171
  172
  173
  174
  175
  176
  177
  178
  179
  180
  181
  182
  183
  184
  185
  186
  187
  188
  189
  190
  191
  192
  193
  194
  195
  196
  197
  198
  199
  200
  201
  202
  203
  204
  205
  206
  207
  208
  209
  210
  211
  212
  213
  214
  215
  216
  217
  218
  219
  220
  221
  222
  223
  224
  225
  226
  227
  228
  229
  230
  231
  232
  233
  234
  235
  236
  237
  238
  239
  240
  241
  242
  243
  244
  245
  246
  247
  248
  249
  250
  251
  252
  253
  254
  255
  256
  257
  258
  259
  260
  261
  262
  263
  264
  265
  266
  267
  268
  269
  270
  271
  272
  273
  274
  275
  276
  277
  278
  279
  280
  281
  282
  283
  284
  285
  286
  287
  288
  289
  290
  291
  292
  293
  294
  295
  296
  297
  298
  299
  300
  301
  302
  303
  304
  305
  306
  307
  308
  309
  310
  311
  312
  313
  314
  315
  316
  317
  318
  319
  320
  321
  322
  323
  324
  325
  326
  327
  328
  329
  330
  331
  332
  333
  334
  335
  336
  337
  338
  339
  340
  341
  342
  343
  344
  345
  346
  347
  348
  349
  350
  351
  352
  353
  354
  355
  356
  357
  358
  359
  360
  361
  362
  363
  364
  365
  366
  367
  368
  369
  370
  371
  372
  373
  374
  375
  376
  377
  378
  379
  380
  381
  382
  383
  384
  385
  386
  387
  388
  389
  390
  391
  392
  393
  394
  395
  396
  397
  398
  399
  400
  401
  402
  403
  404
  405
  406
  407
  408
  409
  410
  411
  412
  413
  414
  415
  416
  417
  418
  419
  420
  421
  422
  423
  424
  425
  426
  427
  428
  429
  430
  431
  432
  433
  434
  435
  436
  437
  438
  439
  440
  441
  442
  443
  444
  445
  446
  447
  448
  449
  450
  451
  452
  453
  454
  455
  456
  457
  458
  459
  460
  461
  462
  463
  464
  465
  466
  467
  468
  469
  470
  471
  472
  473
  474
  475
  476
  477
  478
  479
  480
  481
  482
  483
  484
  485
  486
  487
  488
  489
  490
  491
  492
  493
  494
  495
  496
  497
  498
  499
  500
  501
  502
  503
  504
  505
  506
  507
  508
  509
  510
  511
  512
  513
  514
  515
  516
  517
  518
  519
  520
  521
  522
  523
  524
  525
  526
  527
  528
  529
  530
  531
  532
  533
  534
  535
  536
  537
  538
  539
  540
  541
  542
  543
  544
  545
  546
  547
  548
  549
  550
  551
  552
  553
  554
  555
  556
  557
  558
  559
  560
  561
  562
  563
  564
  565
  566
  567
  568
  569
  570
  571
  572
  573
  574
  575
  576
  577
  578
  579
  580
  581
  582
  583
  584
  585
  586
  587
  588
  589
  590
  591
  592
  593
  594
  595
  596
  597
  598
  599
  600
  601
  602
  603
  604
  605
  606
  607
  608
  609
  610
  611
  612
  613
  614
  615
  616
  617
  618
  619
  620
  621
  622
  623
  624
  625
  626
  627
  628
  629
  630
  631
  632
  633
  634
  635
  636
  637
  638
  639
  640
  641
  642
  643
  644
  645
  646
  647
  648
  649
  650
  651
  652
  653
  654
  655
  656
  657
  658
  659
  660
  661
  662
  663
  664
  665
  666
  667
  668
  669
  670
  671
  672
  673
  674
  675
  676
  677
  678
  679
  680
  681
  682
  683
  684
  685
  686
  687
  688
  689
  690
  691
  692
  693
  694
  695
  696
  697
  698
  699
  700
  701
  702
  703
  704
  705
  706
  707
  708
  709
  710
  711
  712
  713
  714
  715
  716
  717
  718
  719
  720
  721
  722
  723
  724
  725
  726
  727
  728
  729
  730
  731
  732
  733
  734
  735
  736
  737
  738
  739
  740
  741
  742
  743
  744
  745
  746
  747
  748
  749
  750
  751
  752
  753
  754
  755
  756
  757
  758
  759
  760
  761
  762
  763
  764
  765
  766
  767
  768
  769
  770
  771
  772
  773
  774
  775
  776
  777
  778
  779
  780
  781
  782
  783
  784
  785
  786
  787
  788
  789
  790
  791
  792
  793
  794
  795
  796
  797
  798
  799
  800
  801
  802
  803
  804
  805
  806
  807
  808
  809
  810
  811
  812
  813
  814
  815
  816
  817
  818
  819
  820
  821
  822
  823
  824
  825
  826
  827
  828
  829
  830
  831
  832
  833
  834
  835
  836
  837
  838
  839
  840
  841
  842
  843
  844
  845
  846
  847
  848
  849
  850
  851
  852
  853
  854
  855
  856
  857
  858
  859
  860
  861
  862
  863
  864
  865
  866
  867
  868
  869
  870
  871
  872
  873
  874
  875
  876
  877
  878
  879
  880
  881
  882
  883
  884
  885
  886
  887
  888
  889
  890
  891
  892
  893
  894
  895
  896
  897
  898
  899
  900
  901
  902
  903
  904
  905
  906
  907
  908
  909
  910
  911
  912
  913
  914
  915
  916
  917
  918
  919
  920
  921
  922
  923
  924
  925
  926
  927
  928
  929
  930
  931
  932
  933
  934
  935
  936
  937
  938
  939
  940
  941
  942
  943
  944
  945
  946
  947
  948
  949
  950
  951
  952
  953
  954
  955
  956
  957
  958
  959
  960
  961
  962
  963
  964
  965
  966
  967
  968
  969
  970
  971
  972
  973
  974
  975
  976
  977
  978
  979
  980
  981
  982
  983
  984
  985
  986
  987
  988
  989
  990
  991
  992
  993
  994
  995
  996
  997
  998
  999
  1000
  1001
  1002
  1003
  1004
  1005
  1006
  1007
  1008
  1009
  1010
  1011
  1012
  1013
  1014
  1015
  1016
  1017
  1018
  1019
  1020
  1021
  1022
  1023
  1024
  1025
  1026
  1027
  1028
  1029
  1030
  1031
  1032
  1033
  1034
  1035
  1036
  1037
  1038
  1039
  1040
  1041
  1042
  1043
  1044
  1045
  1046
  1047
  1048
  1049
  1050
  1051
  1052
  1053
  1054
  1055
  1056
  1057
  1058
  1059
  1060
  1061
  1062
  1063
  1064
  1065
  1066
  1067
  1068
  1069
  1070
  1071
  1072
  1073
  1074
  1075
  1076
  1077
  1078
  1079
  1080
  1081
  1082
  1083
  1084
  1085
  1086
  1087
  1088
  1089
  1090
  1091
  1092
  1093
  1094
  1095
  1096
  1097
  1098
  1099
  1100
  1101
  1102
  1103
  1104
  1105
  1106
  1107
  1108
  1109
  1110
  1111
  1112
  1113
  1114
  1115
  1116
  1117
  1118
  1119
  1120
  1121
  1122
  1123
  1124
  1125
  1126
  1127
  1128
  1129
  1130
  1131
  1132
  1133
  1134
  1135
  1136
  1137
  1138
  1139
  1140
  1141
  1142
  1143
  1144
  1145
  1146
  1147
  1148
  1149
  1150
  1151
  1152
  1153
  1154
  1155
  1156
  1157
  1158
  1159
  1160
  1161
  1162
  1163
  1164
  1165
  1166
  1167
  1168
  1169
  1170
  1171
  1172
  1173
  1174
  1175
  1176
  1177
  1178
  1179
  1180
  1181
  1182
  1183
  1184
  1185
  1186
  1187
  1188
  1189
  1190
  1191
  1192
  1193
  1194
  1195
  1196
  1197
  1198
  1199
  1200
  1201
  1202
  1203
  1204
  1205
  1206
  1207
  1208
  1209
  1210
  1211
  1212
  1213
  1214
  1215
  1216
  1217
  1218
  1219
  1220
  1221
  1222
  1223
  1224
  1225
  1226
  1227
  1228
  1229
  1230
  1231
  1232
  1233
  1234
  1235
  1236
  1237
  1238
  1239
  1240
  1241
  1242
  1243
  1244
  1245
  1246
  1247
  1248
  1249
  1250
  1251
  1252
  1253
  1254
  1255
  1256
  1257
  1258
  1259
  1260
  1261
  1262
  1263
  1264
  1265
  1266
  1267
  1268
  1269
  1270
  1271
  1272
  1273
  1274
  1275
  1276
  1277
  1278
  1279
  1280
  1281
  1282
  1283
  1284
  1285
  1286
  1287
  1288
  1289
  1290
  1291
  1292
  1293
  1294
  1295
  1296
  1297
  1298
  1299
  1300
  1301
  1302
  1303
  1304
  1305
  1306
  1307
  1308
  1309
  1310
  1311
  1312
  1313
  1314
  1315
  1316
  1317
  1318
  1319
  1320
  1321
  1322
  1323
  1324
  1325
  1326
  1327
  1328
  1329
  1330
  1331
  1332
  1333
  1334
  1335
  1336
  1337
  1338
  1339
  1340
  1341
  1342
  1343
  1344
  1345
  1346
  1347
  1348
  1349
  1350
  1351
  1352
  1353
  1354
  1355
  1356
  1357
  1358
  1359
  1360
  1361
  1362
  1363
  1364
  1365
  1366
  1367
  1368
  1369
  1370
  1371
  1372
  1373
  1374
  1375
  1376
  1377
  1378
  1379
  1380
  1381
  1382
  1383
  1384
  1385
  1386
  1387
  1388
  1389
  1390
  1391
  1392
  1393
  1394
  1395
  1396
  1397
  1398
  1399
  1400
  1401
  1402
  1403
  1404
  1405
  1406
  1407
  1408
  1409
  1410
  1411
  1412
  1413
  1414
  1415
  1416
  1417
  1418
  1419
  1420
  1421
  1422
  1423
  1424
  1425
  1426
  1427
  1428
  1429
  1430
  1431
  1432
  1433
  1434
  1435
  1436
  1437
  1438
  1439
  1440
  1441
  1442
  1443
  1444
  1445
  1446
  1447
  1448
  1449
  1450
  1451
  1452
  1453
  1454
  1455
  1456
  1457
  1458
  1459
  1460
  1461
  1462
  1463
  1464
  1465
  1466
  1467
  1468
  1469
  1470
  1471
  1472
  1473
  1474
  1475
  1476
  1477
  1478
  1479
  1480
  1481
  1482
  1483
  1484
  1485
  1486
  1487
  1488
  1489
  1490
  1491
  1492
  1493
  1494
  1495
  1496
  1497
  1498
  1499
  1500
  1501
  1502
  1503
  1504
  1505
  1506
  1507
  1508
  1509
  1510
  1511
  1512
  1513
  1514
  1515
  1516
  1517
  1518
  1519
  1520
  1521
  1522
  1523
  1524
  1525
  1526
  1527
  1528
  1529
  1530
  1531
  1532
  1533
  1534
  1535
  1536
  1537
  1538
  1539
  1540
  1541
  1542
  1543
  1544
  1545
  1546
  1547
  1548
  1549
  1550
  1551
  1552
  1553
  1554
  1555
  1556
  1557
  1558
  1559
  1560
  1561
  1562
  1563
  1564
  1565
  1566
  1567
  1568
  1569
  1570
  1571
  1572
  1573
  1574
  1575
  1576
  1577
  1578
  1579
  1580
  1581
  1582
  1583
  1584
  1585
  1586
  1587
  1588
  1589
  1590
  1591
  1592
  1593
  1594
  1595
  1596
  1597
  1598
  1599
  1600
  1601
  1602
  1603
  1604
  1605
  1606
  1607
  1608
  1609
  1610
  1611
  1612
  1613
  1614
  1615
  1616
  1617
  1618
  1619
  1620
  1621
  1622
  1623
  1624
  1625
  1626
  1627
  1628
  1629
  1630
  1631
  1632
  1633
  1634
  1635
  1636
  1637
  1638
  1639
  1640
  1641
  1642
  1643
  1644
  1645
  1646
  1647
  1648
  1649
  1650
  1651
  1652
  1653
  1654
  1655
  1656
  1657
  1658
  1659
  1660
  1661
  1662
  1663
  1664
  1665
  1666
  1667
  1668
  1669
  1670
  1671
  1672
  1673
  1674
  1675
  1676
  1677
  1678
  1679
  1680
  1681
  1682
  1683
  1684
  1685
  1686
  1687
  1688
  1689
  1690
  1691
  1692
  1693
  1694
  1695
  1696
  1697
  1698
  1699
  1700
  1701
  1702
  1703
  1704
  1705
  1706
  1707
  1708
  1709
  1710
  1711
  1712
  1713
  1714
  1715
  1716
  1717
  1718
  1719
  1720
  1721
  1722
  1723
  1724
  1725
  1726
  1727
  1728
  1729
  1730
  1731
  1732
  1733
  1734
  1735
  1736
  1737
  1738
  1739
  1740
  1741
  1742
  1743
  1744
  1745
  1746
  1747
  1748
  1749
  1750
  1751
  1752
  1753
  1754
  1755
  1756
  1757
  1758
  1759
  1760
  1761
  1762
  1763
  1764
  1765
  1766
  1767
  1768
  1769
  1770
  1771
  1772
  1773
  1774
  1775
  1776
  1777
  1778
  1779
  1780
  1781
  1782
  1783
  1784
  1785
  1786
  1787
  1788
  1789
  1790
  1791
  1792
  1793
  1794
  1795
  1796
  1797
  1798
  1799
  1800
  1801
  1802
  1803
  1804
  1805
  1806
  1807
  1808
  1809
  1810
  1811
  1812
  1813
  1814
  1815
  1816
  1817
  1818
  1819
  1820
  1821
  1822
  1823
  1824
  1825
  1826
  1827
  1828
  1829
  1830
  1831
  1832
  1833
  1834
  1835
  1836
  1837
  1838
  1839
  1840
  1841
  1842
  1843
  1844
  1845
  1846
  1847
  1848
  1849
  1850
  1851
  1852
  1853
  1854
  1855
  1856
  1857
  1858
  1859
  1860
  1861
  1862
  1863
  1864
  1865
  1866
  1867
  1868
  1869
  1870
  1871
  1872
  1873
  1874
  1875
  1876
  1877
  1878
  1879
  1880
  1881
  1882
  1883
  1884
  1885
  1886
  1887
  1888
  1889
  1890
  1891
  1892
  1893
  1894
  1895
  
```

报告要求

- 任务一：完善仿真环境中的**奖励函数**。

[要求] 设计合理的奖励函数，使得强化学习网络能够根据该奖励函数学习到正确的动作

- 任务二：定义**动作空间和观测空间**。

[要求]

1. 完成环境初始化中的observation space和action space的定义
2. 完成_get_obs函数，从环境中得到观测信息。

- 任务三：在任务二的基础上完善**step函数**，根据输入动作下发机械臂控制指令。利用test_gym.py验证动作的可行性。
- 任务四：基于DQN算法，完成**操作策略的训练**，记录训练曲线，并测试效果。

[要求]

1. 基于提供的DQN代码，训练抓取任务的强化学习策略
2. 使用tensorboard添加训练信息记录，并可视化训练数据
3. 测试策略效果，测试20次，记录成功率；并在提交附件中包含视频形式的结果展示
4. 通过可视化的训练数据和策略效果，分析此次训练的结果，并提出未来可能的改进方向

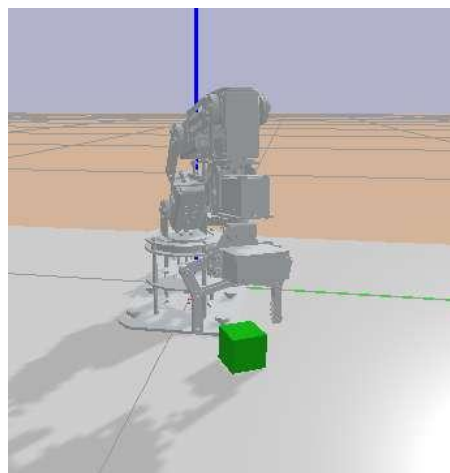
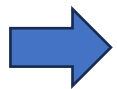
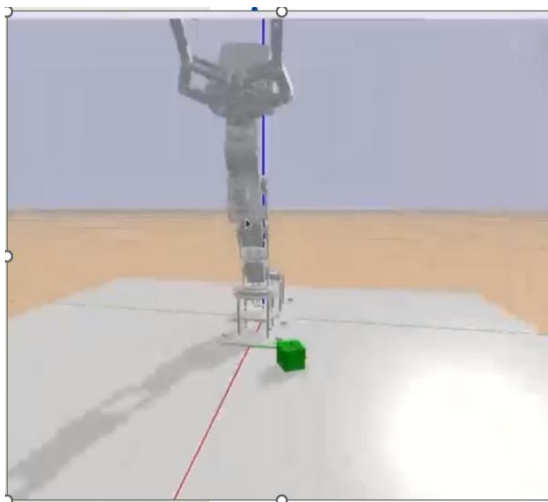
思考：有没有办法可以通过强化学习训练Dofbot这种缺少自由度的机械臂？

答案：

- 确保完成任务的途径点，机械臂末端都是可达的
- 将关节空间作为动作空间，不存在解逆运动学的问题

支持的算法：可用于**连续动作空间**的强化学习算法

- PPO、SAC.... (操作领域用的较多，算法效果好)

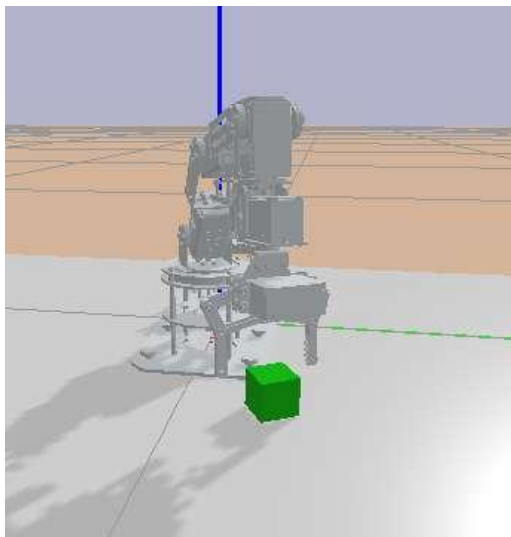
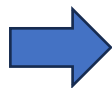
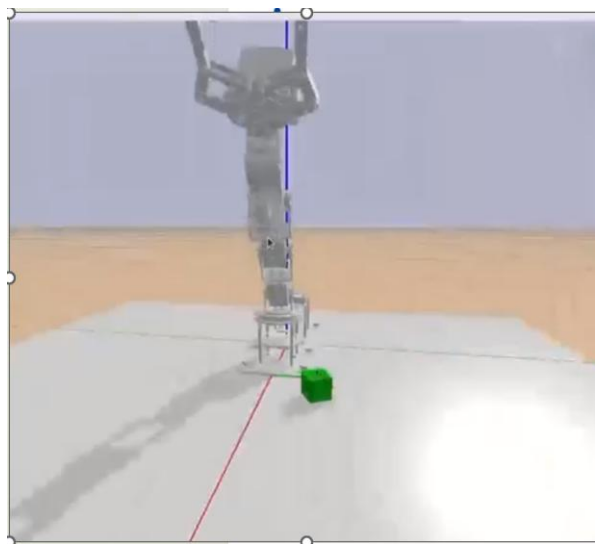


任务介绍

任务目标

使用仿真环境pybullet完善抓取方块任务的仿真环境，并使用强化学习算法（SAC），完成抓取方块任务的操作策略的训练和优化。

为简化任务设置，减少训练时间，本次任务只需要完成靠近目标物体的任务。即机械臂夹爪距离目标物体的距离小于阈值（0.01m）则视为任务成功。



任务环境

智能体：

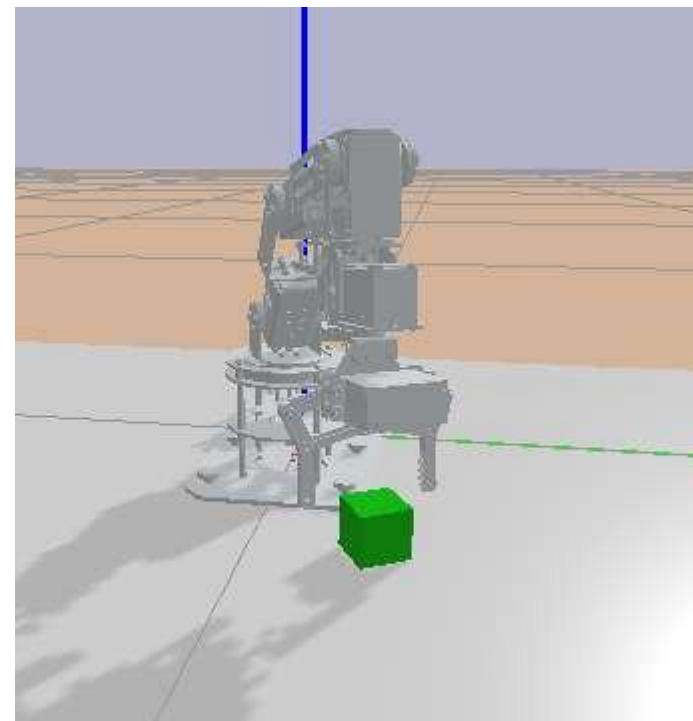
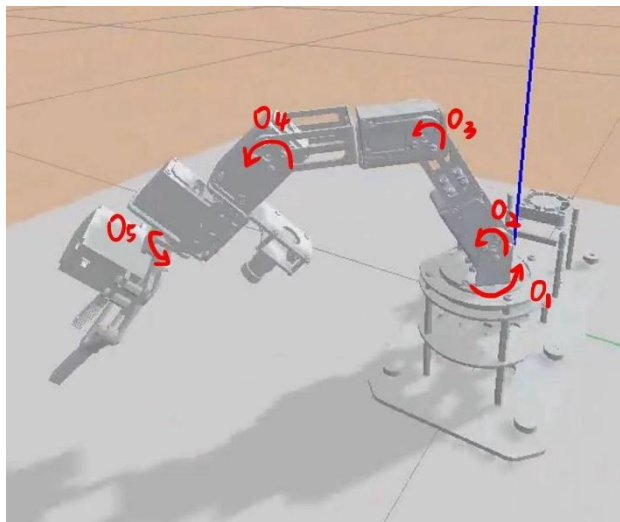
Dofbot机械臂，5自由度。

观察：

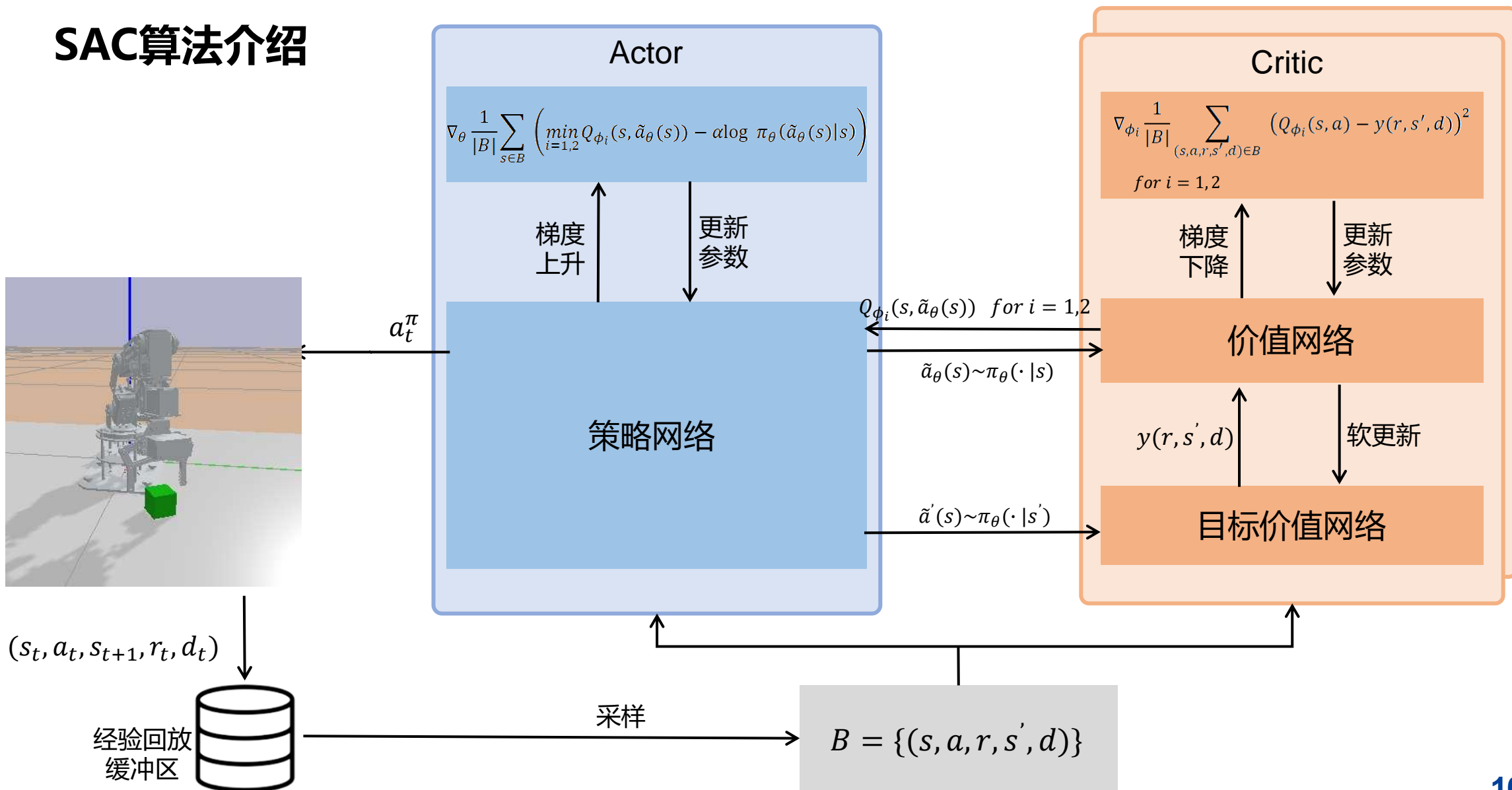
- 机械臂关节位置（6维）：5维机械臂关节+1维夹爪
- 机械臂夹爪位姿：7维（3维位置+4维四元数）
- 物体在机械臂夹爪坐标系下的位姿：7维（3维位置+4维四元数）

动作（连续空间）：

- 机械臂5个关节的变化位置



SAC算法介绍



拓展任务（机械臂强化学习控制部分）

报告要求

- 任务一：完善仿真环境中的**奖励函数**。

[要求] 设计合理的奖励函数，使得强化学习网络能够根据该奖励函数学习到正确的动作

- 任务二：定义**动作空间和观测空间**。

[要求]

1. 完成环境初始化中的observation space和action space的定义
2. 完成_get_obs函数，从环境中得到观测信息。

- 任务三：在任务二的基础上完善**step函数**，根据输入动作下发机械臂控制指令。
- 任务四：通过求解逆运动学，找到逆运动学可解的位置，放置方块。
- 任务五：基于SAC算法，完成**操作策略的训练**，记录训练曲线，并测试效果。

[要求]

1. 基于提供的SAC代码，训练抓取任务的强化学习策略
2. 使用tensorboard添加训练信息记录，并可视化训练数据
3. 测试策略效果，测试50次，记录成功率；并在提交附件中包含视频形式的结果展示
4. 通过可视化的训练数据和策略效果，分析此次训练的结果，并提出未来可能的改进方向

关节角度控制

求解逆运动学

获取关节位置

获取关节速度

```
def joint_control(self, dqpos):
    self.desire_qpos = self.desire_qpos + dqpos
    jointPoses = self.desire_qpos
    for i in range(self.numJoints):
        p.setJointMotorControl2(bodyUniqueId=self.dofbotUid, jointIndex=i, controlMode=p.POSITION_CONTROL,
                                targetPosition=jointPoses[i], targetVelocity=0, force=200,
                                maxVelocity=10.0, positionGain=0.3, velocityGain=1)
    self.jointPositions, self.gripperAngle = self.get_jointPoses()
    self.endEffectorPos, self.endEffectorOrn, self.endEffectorEuler = self.get_pose()
    return self.endEffectorPos, self.endEffectorOrn, self.endEffectorEuler

def setInverseKine(self, pos, orn):
    if orn == None:
        jointPoses = p.calculateInverseKinematics(self.dofbotUid, 4, pos,
                                                  self.ll, self.ul, self.jr, self.rp)
    else:
        jointPoses = p.calculateInverseKinematics(self.dofbotUid, 4, pos, orn,
                                                  self.ll, self.ul, self.jr, self.rp)
    return jointPoses[:self.numJoints], self.gripperAngle

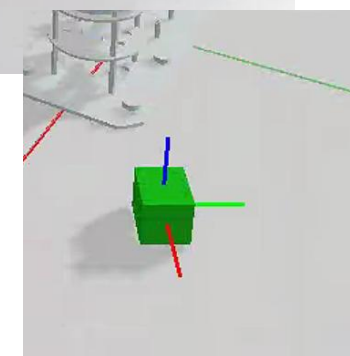
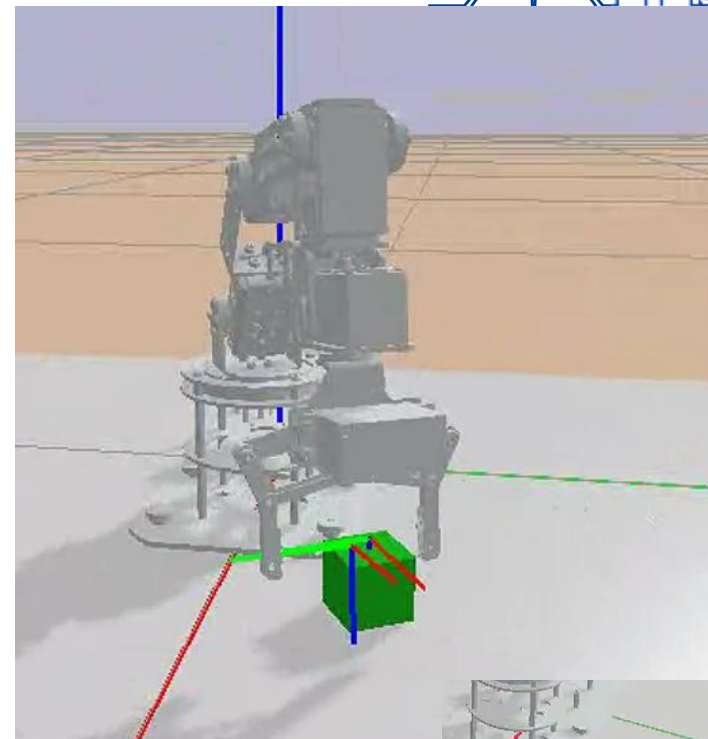
def get_jointPoses(self):
    jointPoses = []
    for i in range(self.numJoints+1):
        state = p.getJointState(self.dofbotUid, i)
        jointPoses.append(state[0])
    return jointPoses[:self.numJoints], self.gripperAngle

def get_qvel(self):
    jointVels = []
    for i in range(self.numJoints+1):
        state = p.getJointState(self.dofbotUid, i)
        jointVels.append(state[1])
    return np.array(jointVels[:self.numJoints])
```

任务一：在代码dofbotGymEnv.py中，完善仿真环境中的**奖励函数**。

提示：可以考虑以下几个方向

- **靠近奖励**：目标方块和夹爪（TCP）的距离越近，奖励越大。
- **姿态奖励**：夹爪姿态与期望抓取姿态的差越小，奖励越大。
- **到达奖励**：当目标方块和夹爪的距离小于目标阈值时，给予奖励。



```
# TODO: design suitable reward function
def _get_reward(self):
    obs = self._get_obs_dict()
    info = self._get_info()

    reward = 0
    return reward
```

[要求]

设计合理的奖励函数，使得强化学习网络能够根据该奖励函数学习到正确的动作

- 任务二：定义动作空间和观测空间。

[要求1] 完成环境初始化中的observation space和action space的定义。

为了训练方便，一般设置动作空间大小为 $[-1, 1]$

```
# TODO: define observation space and action space
# self.observation_space =
# self.action_space =
```

[要求2]完成_get_obs函数，从环境中得到观测信息。

```
# TODO: complete observation
def _get_obs(self):
    Observation = self._panda.getObservation()

    # TODO: add suitable observation items here

    if self.obs_mode == "state_dict":
        self._observation = Observation
        return self._observation
    elif self.obs_mode == "state":
        values = list(Observation.values())
        self._observation = np.concatenate([v if isinstance(v, np.ndarray)
        self._observation = self._observation.astype(np.float32)
        return self._observation
```

获取物体位姿函数：

- getBasePositionAndOrientation(objectUnique Id)：返回位置列表（包含3个浮点数）以及方向列表（包含4个浮点数，按 [x, y, z, w] 顺序排列）。

- 任务三：在任务二的基础上完善**step函数**，根据输入动作下发机械臂控制指令。

提示：注意动作空间的范围，要转换为真实的机械臂关节位置变化量的大小

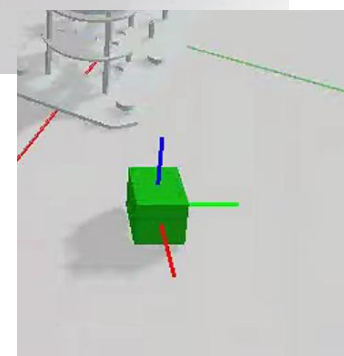
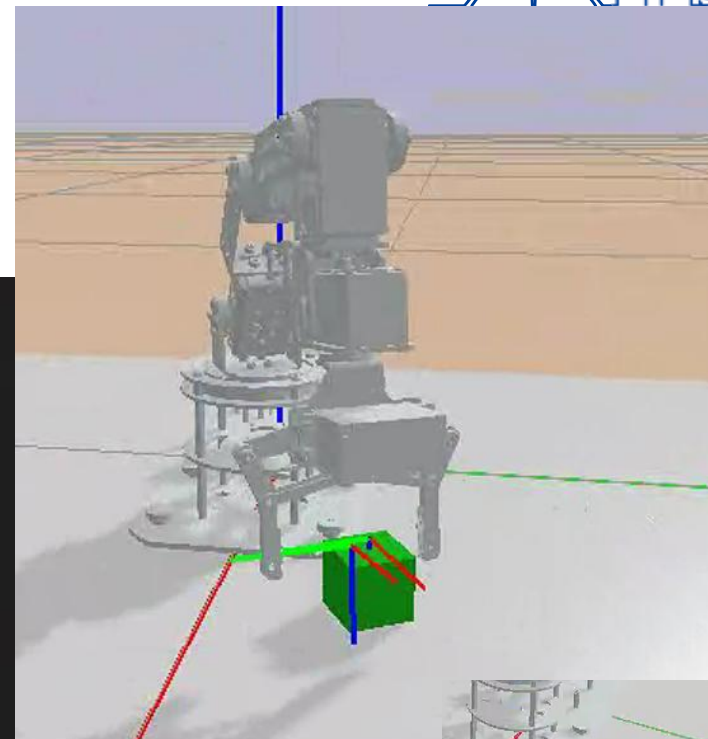
```
def step(self, action):  
    """  
    action - qpos np.array(5), gripper keeps open  
    """  
    # TODO: complete step control of dofbot  
  
    for i in range(self.simuRepeatNum):  
        p.stepSimulation()  
  
    if self.render_mode == "human":  
        time.sleep(self._timeStep)  
    terminated = self._termination()  
    truncated = False  
    self._observation = self._get_obs()  
    reward = self._get_reward()  
    info = self._get_info()  
    return self._observation, reward, terminated, truncated, info
```


- 任务四：通过求解逆运动学，找到逆运动学可解的位姿，放置方块。

```
class Object:
    def __init__(self, urdfPath, block,num):
        self.id = p.loadURDF(urdfPath)
        self.half_height = 0.015 if block else 0.0745
        self.num = num

        self.block = block
    def reset(self):

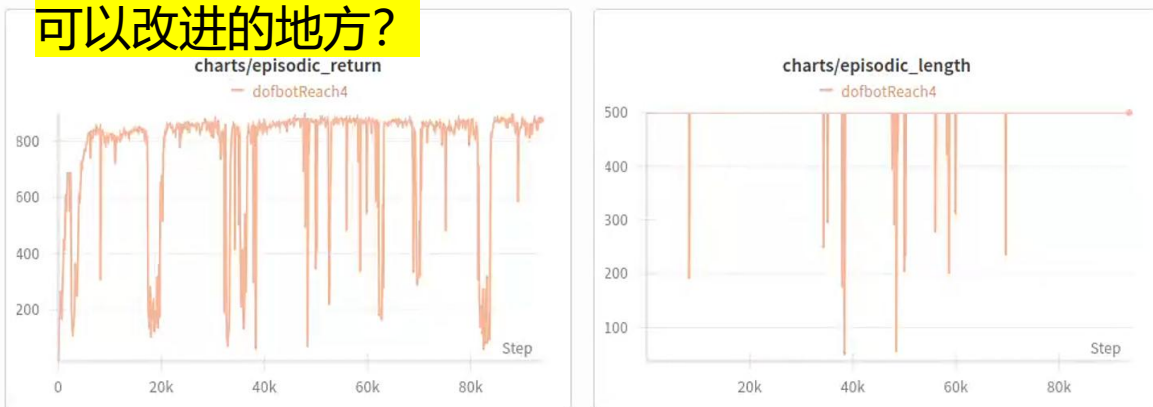
        if self.num==1:
            p.resetBasePositionAndOrientation(self.id,
                                              np.array([ 0.20, 0.1,
                                                         self.half_height]),
                                              p.getQuaternionFromEuler([0, 0,np.pi/6]))
```



任务五：基于SAC算法，完成**操作策略的训练**，记录训练曲线，并测试效果。

训练曲线示例：

思考：为什么训练曲线会有很大的波动？有没有可以改进的地方？



[要求]

1. 基于提供的SAC代码，训练抓取任务的强化学习策略
2. 使用tensorboard添加训练信息记录，并可视化训练数据
3. 测试策略效果并分析原因，在提交附件中包含视频形式的结果展示

训练曲线记录工具：tensorboard

```
pip install tensorboard
```

记录事件：

```
from torch.utils.tensorboard import SummaryWriter

# 初始化SummaryWriter
writer = SummaryWriter('runs/experiment_name')
```

记录信息：

```
writer.add_scalar("losses/td_loss", loss, global_step)
writer.add_scalar("losses/q_values", old_val.mean().item(), global_step)
```

信息可视化：

```
tensorboard --logdir=runs/
```