无人系统设计第四次作业

组员：董云鹏，罗世才，谷金龙，李昱翰，周智涵

任务一：

Observations设计：

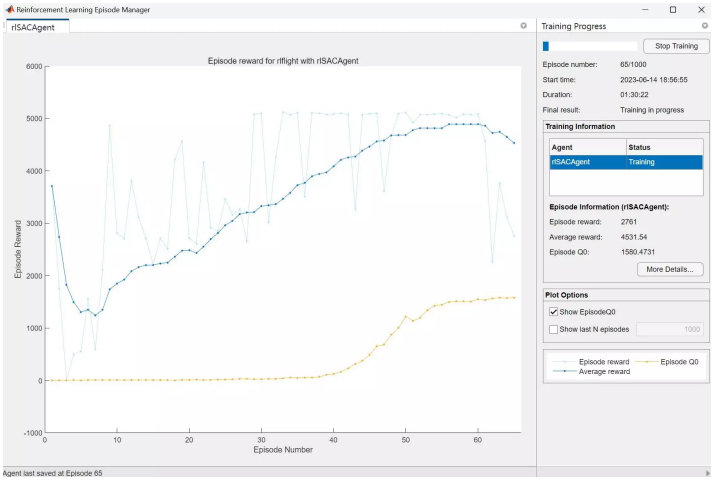
敌机与本机的相对位置rx，ry，rz，本机当前yaw，pitch与目标yaw，pitch（能瞄准敌机的yaw，pitch）的差异yaw_diff，pitch_diff，敌机的yaw_diff和pitch_diff。

Reward设计：

敌机血量被削减时，给予50的奖励。

其他情况下，reward = -0.2 + (abs(pre_pitch_diff) - abs(pitch_diff) + abs(pre_yaw_diff) - abs(yaw_diff)) * 200。其中pre_yaw_diff为前一回合的yaw_diff，pre_pitch_diff为前一回合的pitch_diff。

训练结果截图：



任务二：

Observations设计：

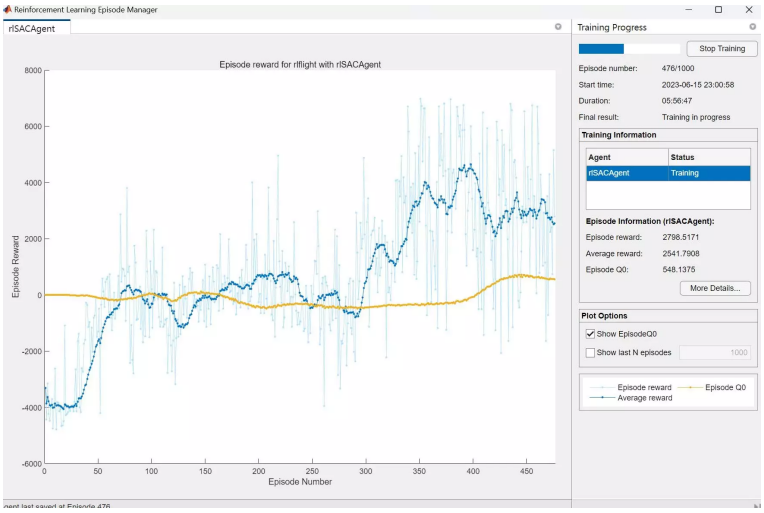
敌机与本机的相对位置rx，ry，rz，本机当前yaw，pitch与目标yaw，pitch（能瞄准敌机的yaw，pitch）的差异yaw_diff，pitch_diff。

Reward设计：

敌机血量被削减时，给予70的奖励。

其他情况下，reward = -0.2 + - 10 * abs(pitch_diff) + 100 * abs(pre_yaw_diff) - 100 * abs(yaw_diff)。其中pre_yaw_diff为前一回合的yaw_diff。

训练结果截图：



任务三：

Observations设计：

敌机与本机的相对位置dx，dy，dz，本机的偏航角yaw，本机当前生命值以及敌机当前生命值，本机的坐标x，y，z。

action只是用了油门以及yaw两个控制量。

Reward设计：

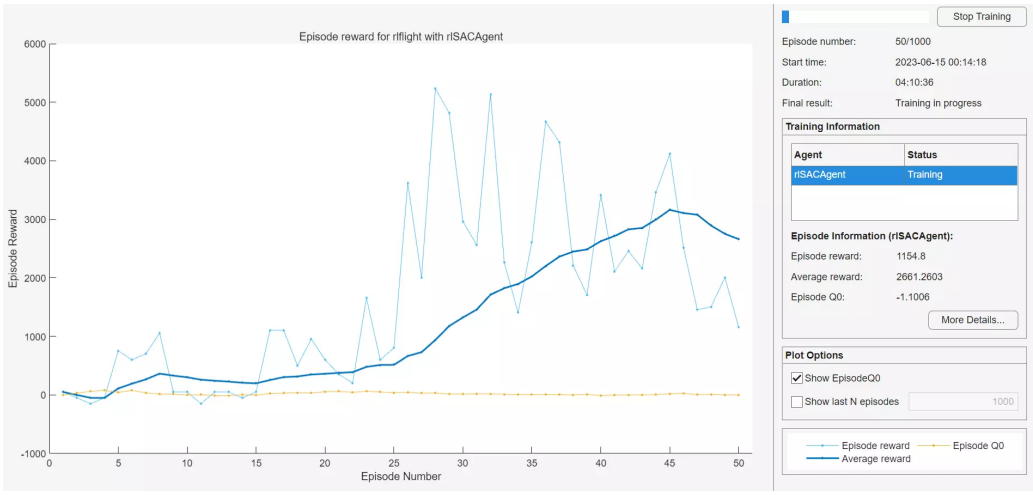
当敌机被成功攻击之后，reward为常量500。

如果没有被攻击，则综合考虑偏航角yaw，本机血量以及敌机血量设计reward如下：

```
reward = -0.2 + 100 * (abs(yaw_diff) - abs(pre_yaw_diff)) + 100 * (pre_enemy_hp - enemy_hp) - 50 * (pre_self_hp - self_hp);
```

训练结果截图：

在以上环境中，训练50轮，每轮1500回合，得到reward曲线如下：



任务四：

修改的任务：

修改了任务2中envs.m文件。

修改的内容：

133行MiniBatchSize从原本的512改为1024。

135行DiscountFactor从原来的0.999变成0.9999。

37行和39行从原来的128变成256。

52行和54行从原来的128变成256。

修改的原因：

训练时观察到，即使已经经过多轮训练，战机有时仍会选择远离正确角度的转向。认为这有可能是观察到的过去经验不足，导致战机在错误的行为中进行选择。修改miniBatchSize希望通过增大对过去经验的采样数，增大观察到正确行为的概率。

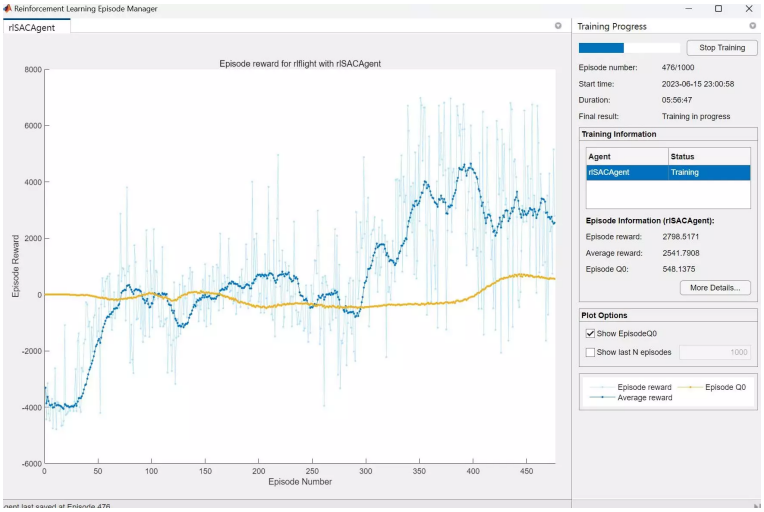
此外，战机有时会在与敌机相反的方向来回摆荡，认为这可能是战机在试图获取势能函数的奖励，而忽视了瞄准敌机后可以获得的大额奖励。因此削减了对于未来奖励的折扣，希望战机能考虑到将来瞄准敌机的大额奖励，从而更快转向到正确位置。

即使限制到了-1到1的范围内，由于战机的各个属性实质上是连续空间中的值，action和observation的空间仍然非常大。增加了网络每层输出的神经元数量，希望网络能提取到更多特征，从而更正确地判断每个行为的奖励。

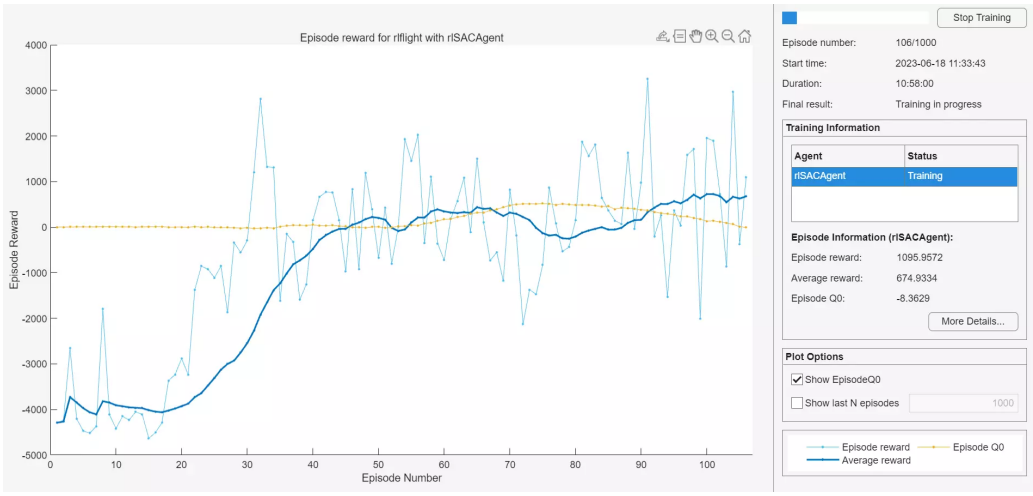
从下面的对比图可以看出，修改之后的版本更快的达到了0以上的average reward，并且reward上升曲线更加平滑。

修改后的变化：

原来的任务二截图：



修改之后的任务二截图：



任务五：

实现的目标：

油门、俯仰、偏航三个控制量击落远距离固定轨迹靶机

Observations设计：

敌机与本机的血量和相对位置rx，ry，rz，本机当前yaw，pitch与目标yaw，pitch（能瞄准敌机的yaw，pitch）的差异yaw_diff和pitch_diff，以及敌机的yaw_diff和pitch_diff。

Reward设计：

Reward分为多组，分别为血量奖励和惩罚，距离奖励，俯仰角奖励，偏航角奖励和角度惩罚。其中敌机血量减少得到奖励70，我机血量减少得到惩罚40。

距离奖励 =（上一回合到敌机距离 - 这一回合到敌机距离）* 0.1，若为负值则乘以2

俯仰角奖励 = - 10 * abs(pitch_diff)，若为负值则乘以2

偏航角奖励 = 100 * abs(pre_yaw_diff) - 100 * abs(yaw_diff)，若为负值则乘以2

角度惩罚 = - 5 * abs(enemy_pitch_diff) + 50 * abs(pre_enemy_yaw_diff) - 50 * abs(enemy_yaw_diff)，若为正值则乘以2

控制惩罚与奖励不对等，减少战机通过反复调整姿态获取奖励的情况，并且攻击到敌机的奖励大于回避攻击的奖励，鼓励战机优先考虑击中敌机。

训练结果截图：