1. 捕获视频只能捕获那一个episode，而这个episode可能不能表示最近的平均水平，最好是多捕获几个。
2. 算的太慢，cpu没用满，可能要用envVector不知道能不能加速。
3. 什么时候停止有点问题。现在6个小时才算6，7w，100w得100个小时了。

跨的自由度不够无法抬脚，没有motion ref全部探索动作空间但是风险太高往往无法进行，还是得需要ref。

Ref motion：可以用urdf得模型然后用ros去控制它去走，来得到ref motion。当务之急是先把sim搞好，把ref motion搞出来。

确定性策略本质上动作空间没变。

只是原地抖动reward都这么高，reward也有问题。