# 2025-02-28

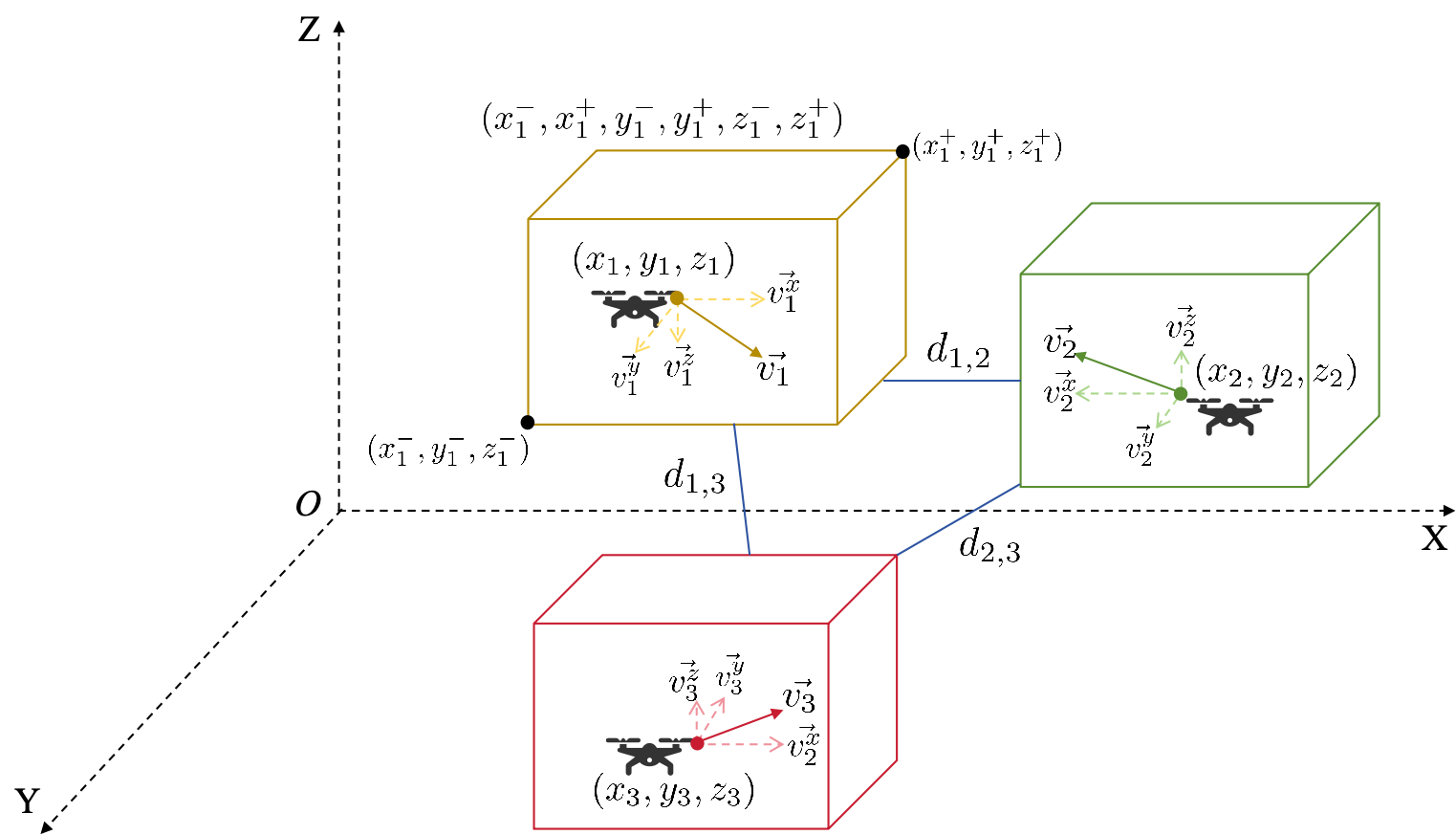
## 目标

在无人机场景中，在每个时刻下计算每个无人机的安全空间，希望在安全空间尽量大的情况下保证无人机自身飞行的安全。

如图空间直角坐标系XOY，当前有n个无人机，每个无人机的坐标(x\_i,y\_i,z\_i)，如下图，各自对应的矩形框为实时分配给无人机的安全空间，满足约束：

1. 无人机各自的安全空间为当前独享，不与其他无人机产生交叉空即冲突；

(2) 在保障不产生冲突的情况下，应当给每个无人机分配尽可能大的空间，以便无人机能尽快完成各自的飞行任务。

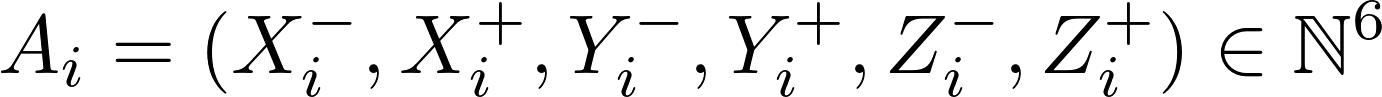


## 目标使用算法

MAPPO算法，多智能体深度强化学习算法，每个智能体代表一个无人机，当前飞行策略是直线飞行，智能体共享一个critic策略网络，全局共享奖励，对于任意一个agent 的参数设置如下：

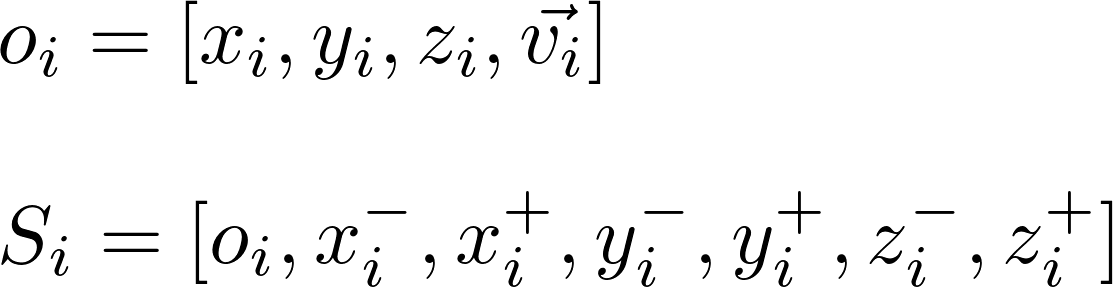
* action

每个agent的动作空间，表示无人机分别在坐标系的六个方向的可移动空间范围，每个方向的取值为自然数(由于每个方向标记+和-，所以取值为自然数即可)，共同组成最终为立方体形状的安全空间。



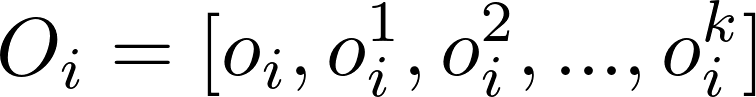
* state

其中表示每个无人机的基本信息元祖，包含其坐标位置和速度(矢量)信息 对于每个agent的状态包含无人机当前的基本信息和安全空间的实际大小



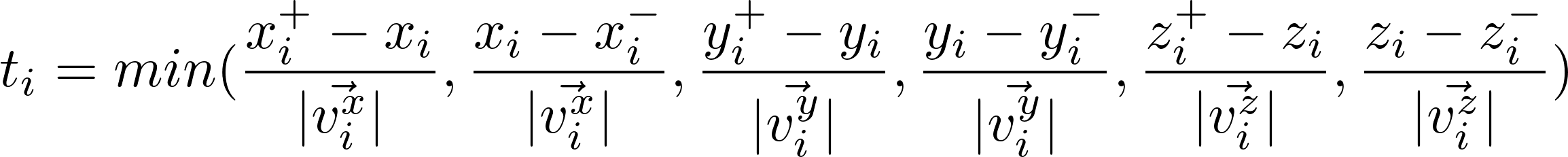
* observation

每个无人机的观测信息为其本身和距离最近的个无人机信息，为训练方便暂时取k=5



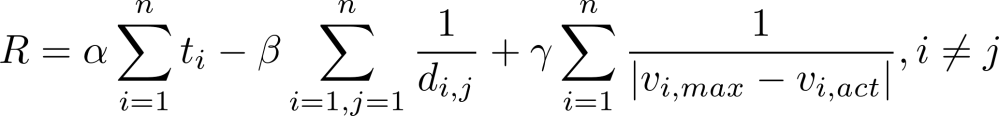
* 奖励，全局共享奖励函数R

其中代表无人机 i 在自己的安全区域，离四面边界的飞行时间，取最小值表示安全时间，表示在这个时间内无人机绝不会突破自己的安全空间，用以满足目标中的约束(2)。



奖励函数R的最终定义，其中中 表示无人机 i 和无人机 j 的安全区域边界之间的最小距离，奖励整体分为三部分，分别为:

1. 无人机的安全时间尽量大，表示权重参数（实际大小可调）；
2. 不同的安全区域之间的距离尽量的大，β表示权重参数（实际大小可调）；
3. 发挥飞行最大性能，是无人机的最大速度，表示无人机当前的实际速度的标量值，应该尽量接近，表示权重参数（实际大小可调）。



* Optimization

1. 当某个agent i从训练流程中移出时：
   1. obs设置成一个对当前agent完全没有威胁信息（补0）
   2. 计算reward时，对reward增加一个过滤机制，把填0的智能体过滤掉