Commercial aircraft control using Deep reinforcement learning

AU332 project presentation

Group member:

Xudong Liu
Lingjia Meng
Buqing Nie

01

Introduction to Flight Gear

Introduction to Flight Gear

How to get data?

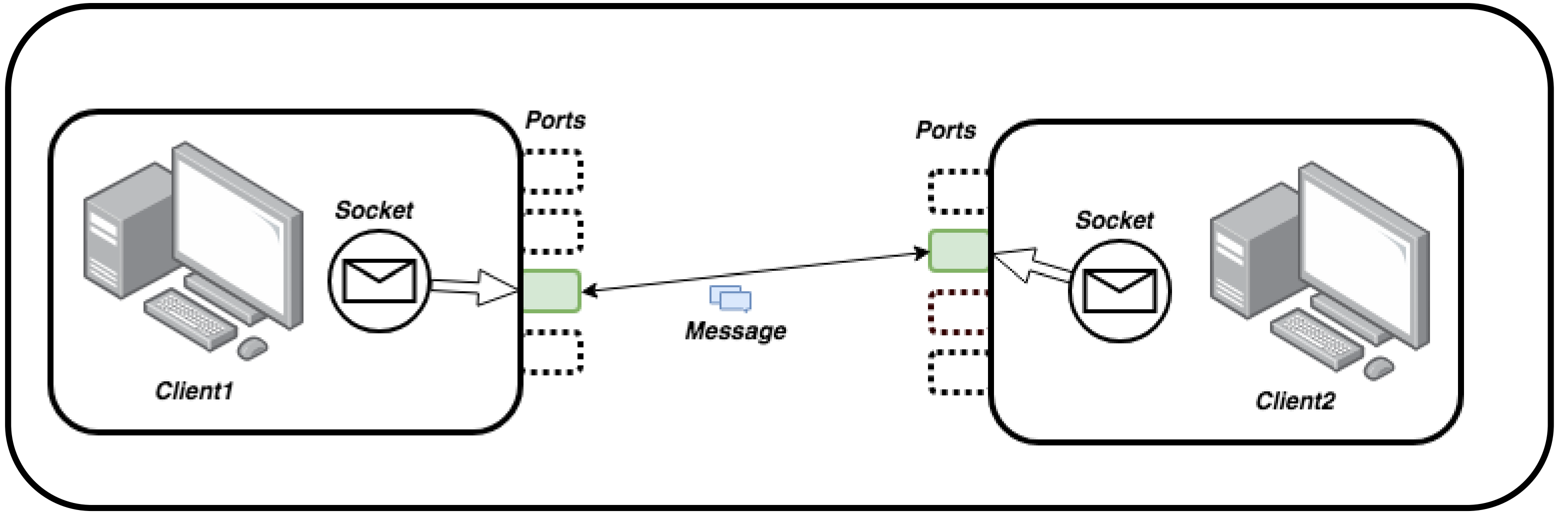
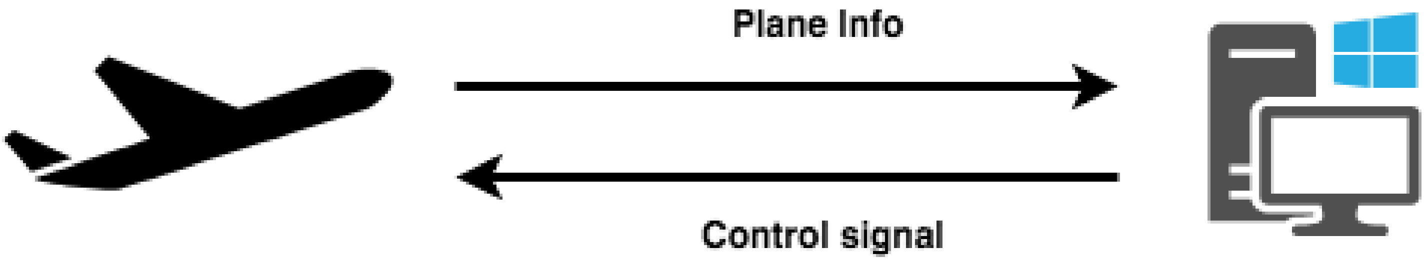
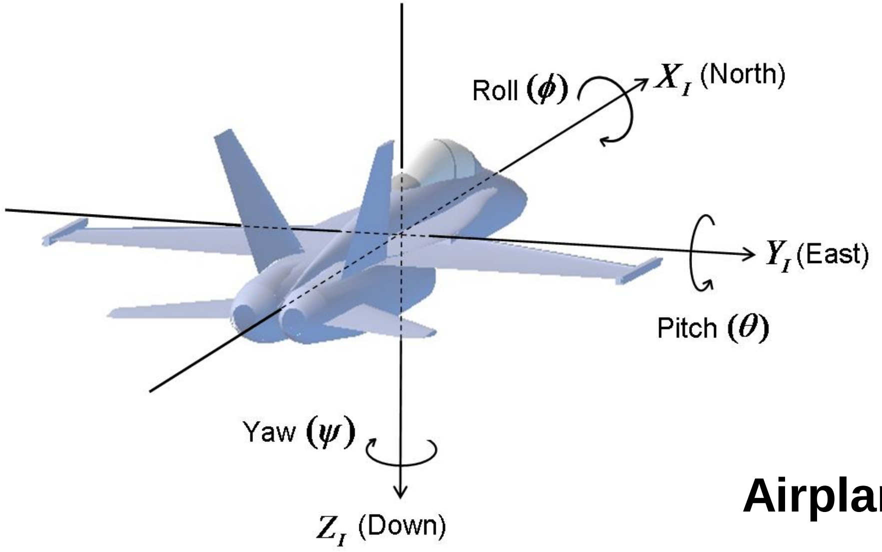


Fig: socket communication figure



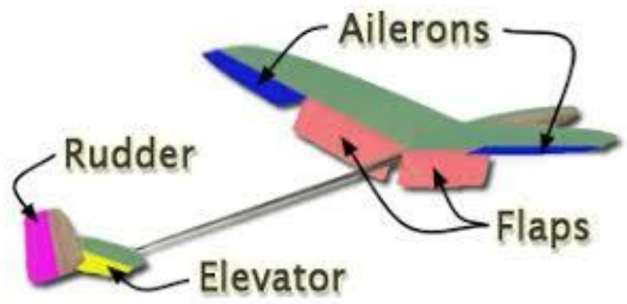
02

Problem formulation

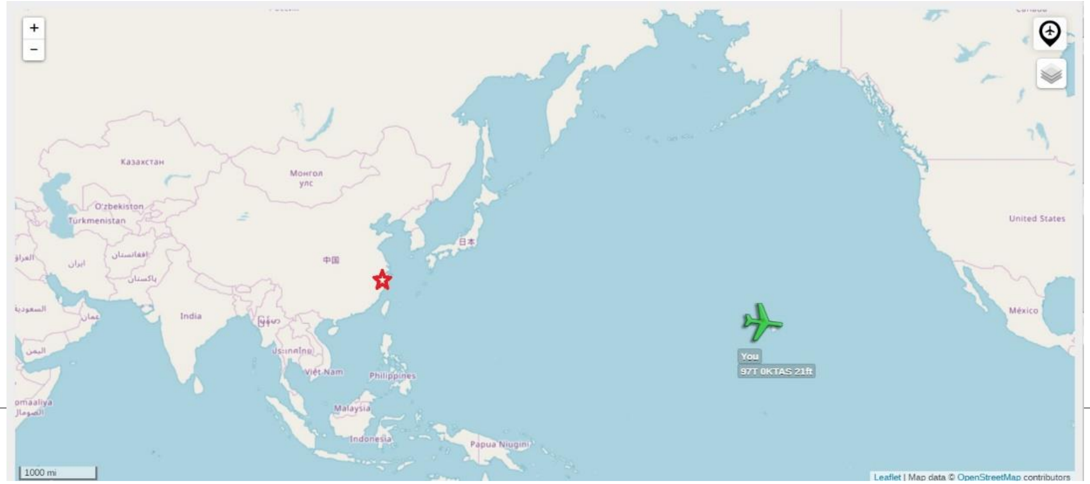


Airplane state

throttles

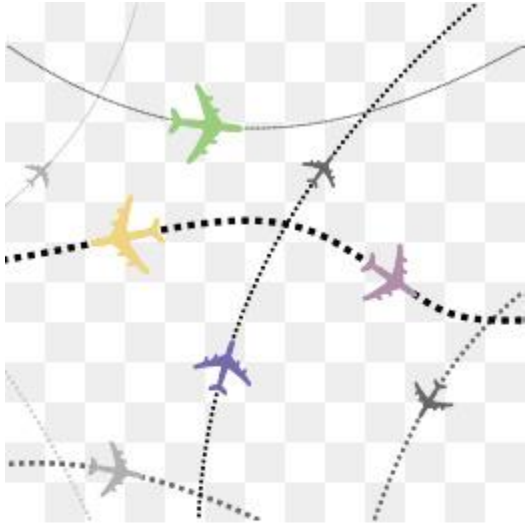


What do we need to control

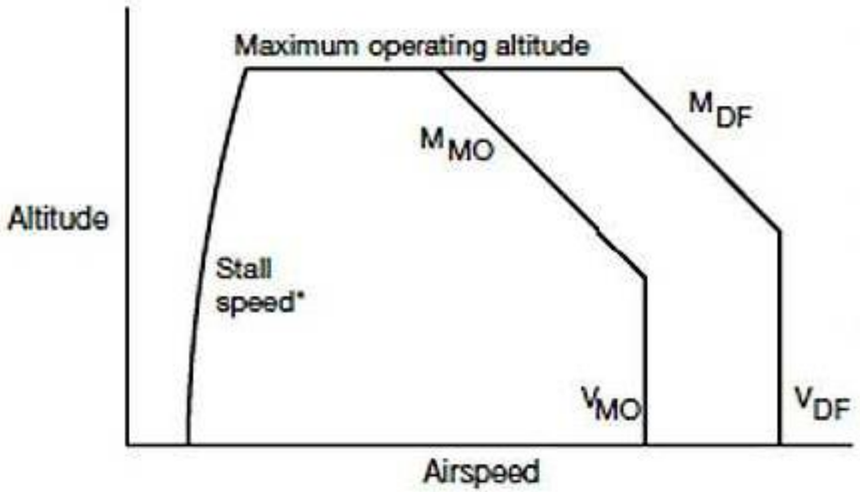


Altitude
latitude, longitude

Goal — flight route and safe flight



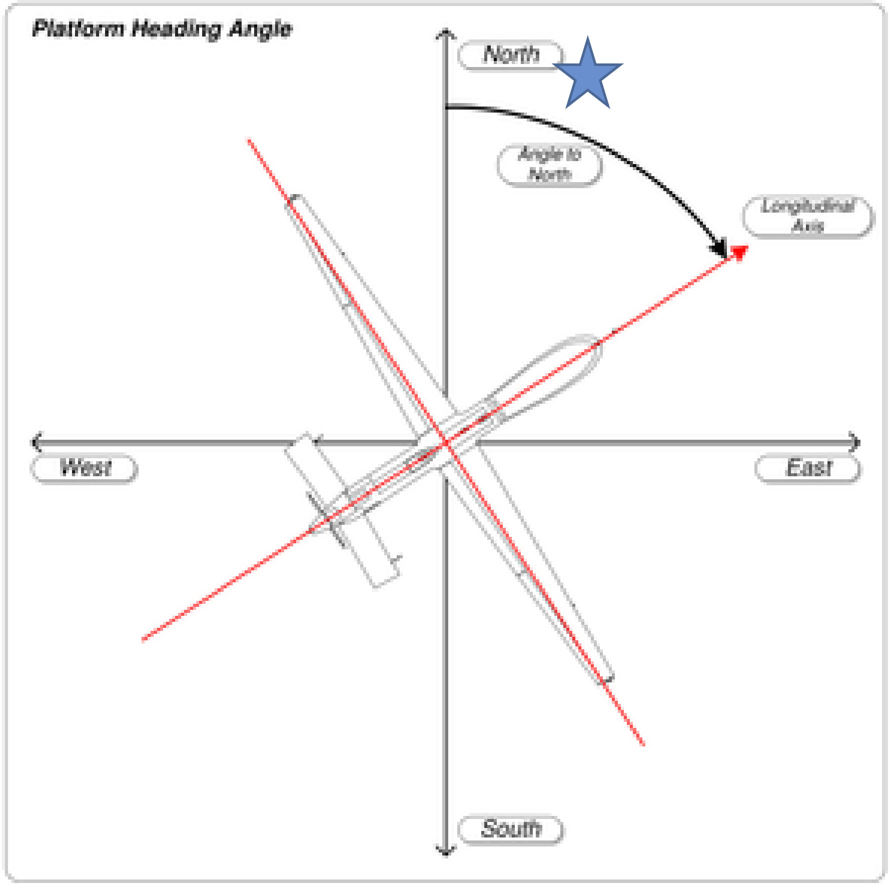
Altitude range
trajectory



Flight envelop
altitude range
speed

How to represent a state
given current information of plane
and target， required range of altitude

Absolute position?



State (Altitude, dist_to_tar, head_offset, angles, vels, accela_vels)

Reward

Different Stages:

Take-off

Flying stably

From dangerous states

……

$$R_{alti} = \omega_1 * |curr\ alti\ - alti\ range| - \omega_2\ * up\_v\_exceed$$

$$R_{dist} = \omega_3 * dist$$

$$R_{heading} = \omega_4 * |head\_offset|$$

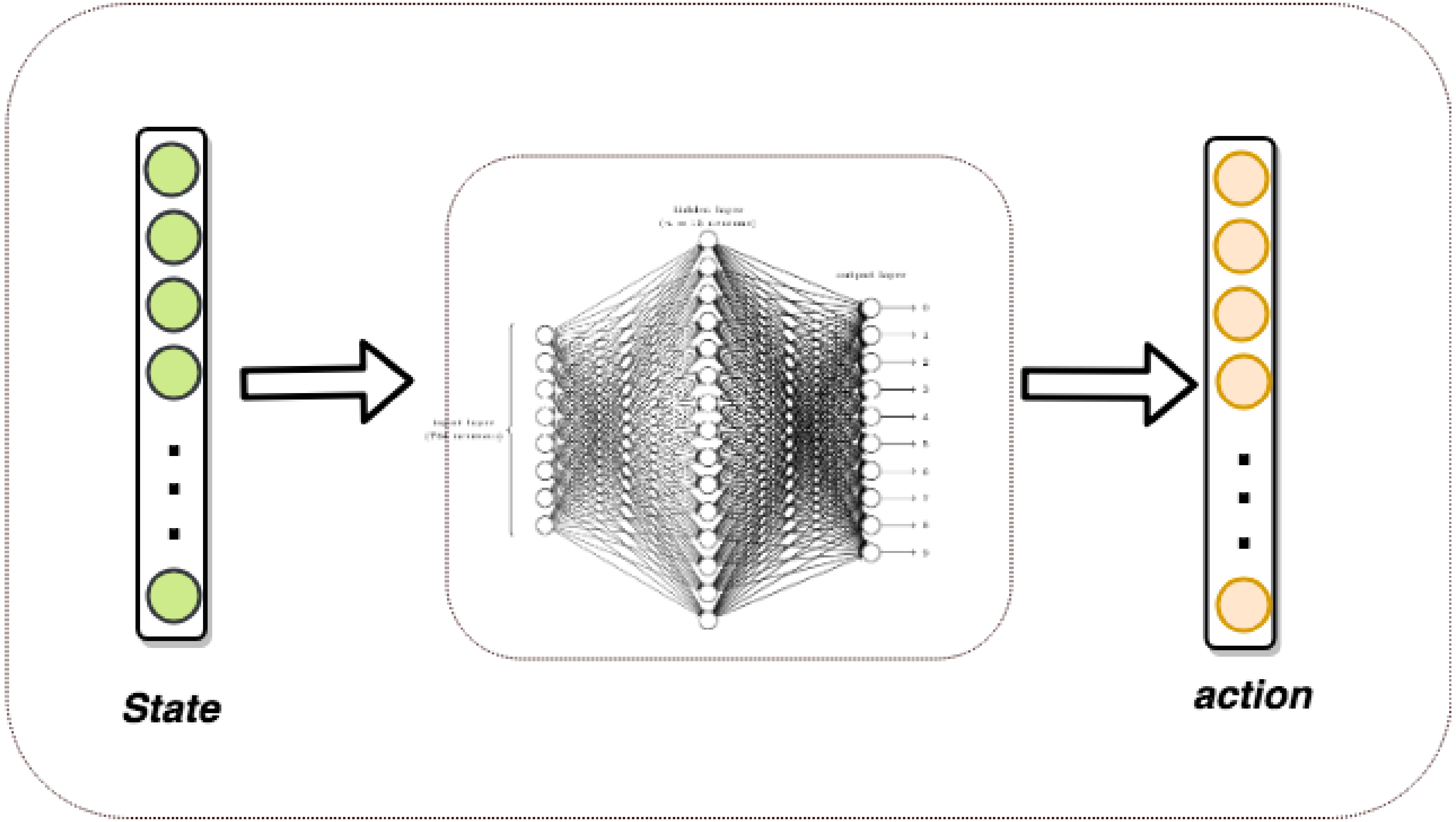
$$R_{stable} = \omega_5 * |\Delta angle| + \omega_6 |\Delta vel|$$

$$R_{danger} = -\omega_7 * e^{|exceed\_speed|}$$

Combine them , will there be a problem ?

$$R = \cdots + (if\ not\ in\ danger) * (R_{stable} + R_{dist} + R_{heading}) + \cdots$$

State -> action



The model we tried

03

DQN & PPO2

DQN algorithm

Q learning

- Replace the Q-table with the neural network
- experience replay for repeated learning
- Q-target mechanism to disrupt correlation

DQN

Advantage *of DQN:*

- Suitable for high dimensional continuous state space situation
  - The state of the plane composite high dimensional continuous state space
- High data utilization for our project
  - The experience replay is used for repeated learning
  - Difficulty for data acquisition

DQN algorithm

Q learning

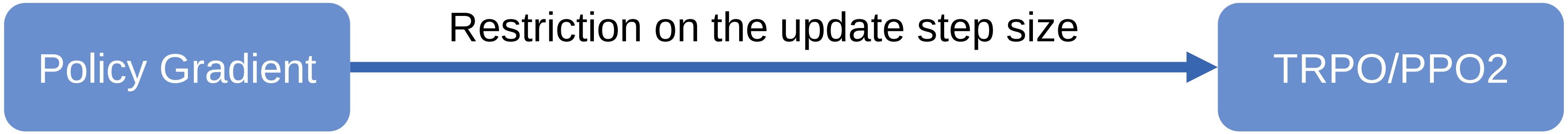
- Replace the Q-table with the neural network
- experience replay for repeated learning
- Q-target mechanism to disrupt correlation

DQN

***Disadvantage** of DQN:*

- The action space must be low dimensional and discrete
  - The output action is discrete, i.e. we must discretize the action space
- Convergence is slow in our project
  - Discretization makes the action space too big, the convergence is too slow in our project.

PPO2(Proximal Policy Optimization) algorithm



Advantage *of PPO2:*

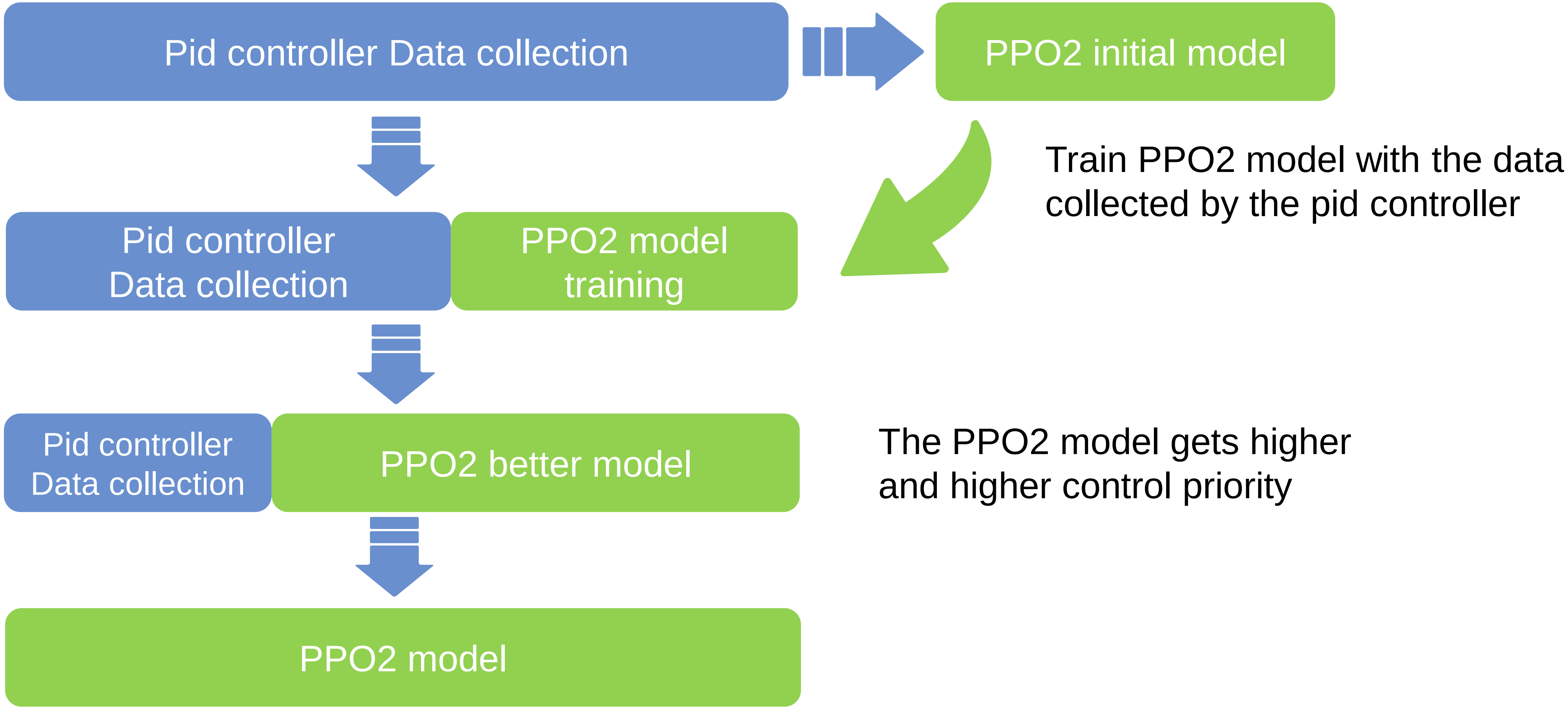
- ➢ Policy-based algorithm, suitable for high dimensional continuous state space & action space
    - The states and actions of the plane are continuous, and discretization is harmful to converge

What we are doing

04

PPO2 + pid controller

What we are doing



What we are doing

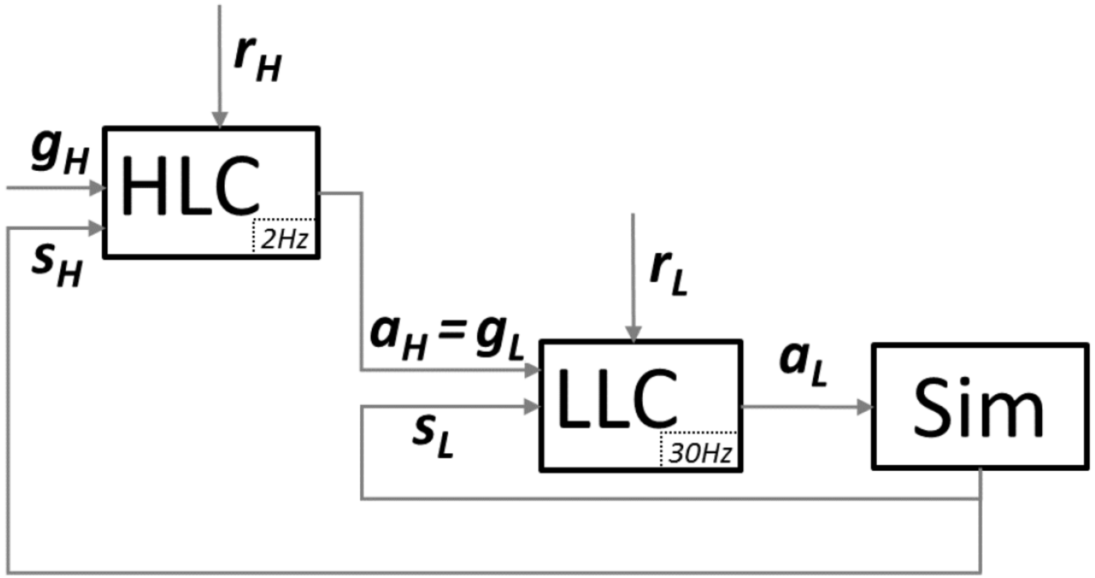


Fig: a two level controller architecture

- HLC: Flight path control
- LLC: Flight posture control

Challenge

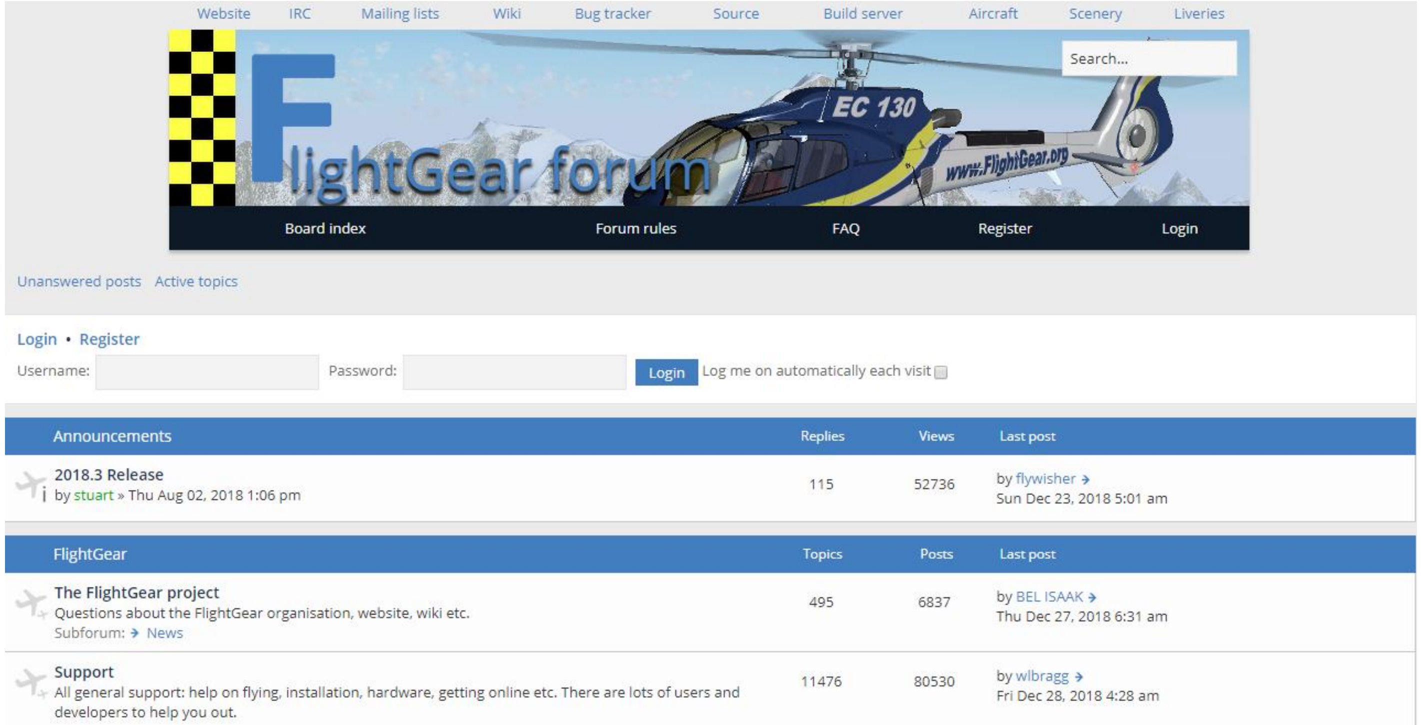
04

Supporting text here.

When you copy & paste, choose "keep text only" option.

Data collection

Quality Diversity （perform randomly）



Imitation ?

S -> a -> s -> a ->… sampling frequency

Trajectory planning

Given a long journey, how to do local planning?

Not only algorithms

Redefine the state and reward

self restriction

Interaction with the software

Insight needs efforts

When training : Reset takes long time

Thank for your listening

2018/12/28