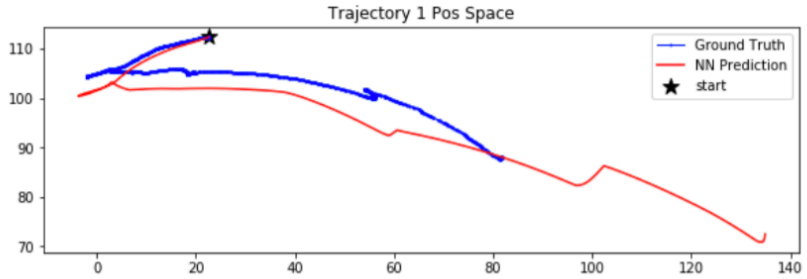
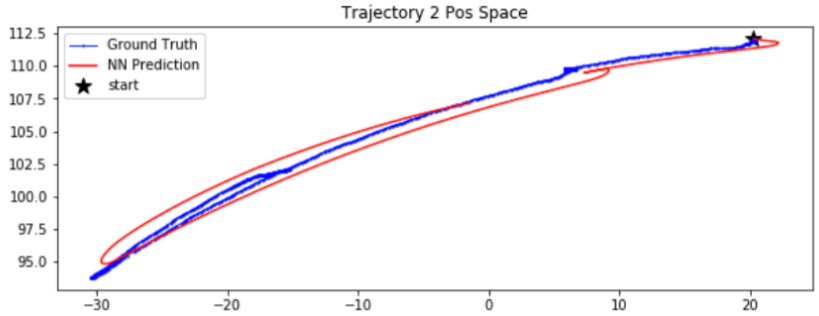
Meeting
10/08/2020

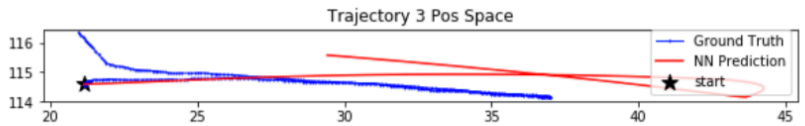
Shuo Zhang

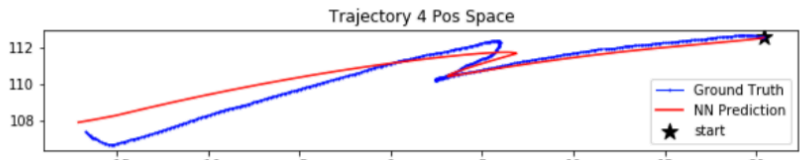
Real Hand

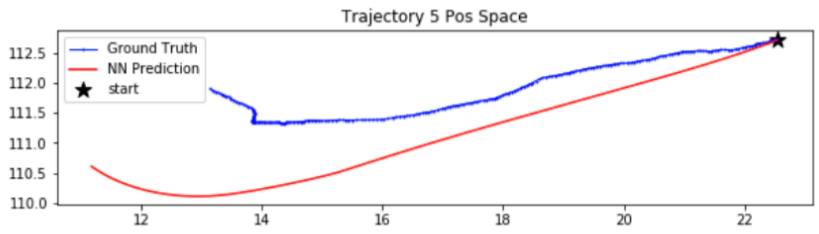
- Fixed all issues of marker tracking and data collection with Mridul
- Got 20 trajectories from Mridul to recalibrate and test the validity of previously trained NN
- Previous NN dynamics works not bad generally, though open-loop rollout failure might happen.
- LQR closed-loop control or Reinforcement learning /AIP are likely to help rollout.

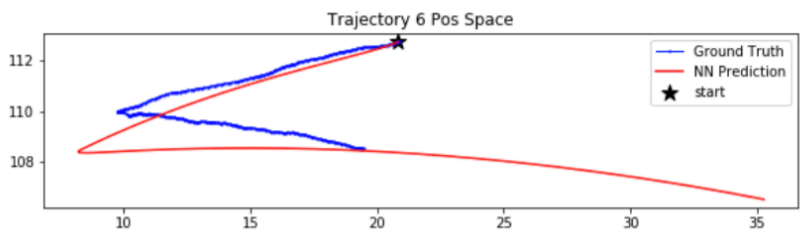

Trajectory 1 Pos Space
Ground Truth
NN Prediction
start

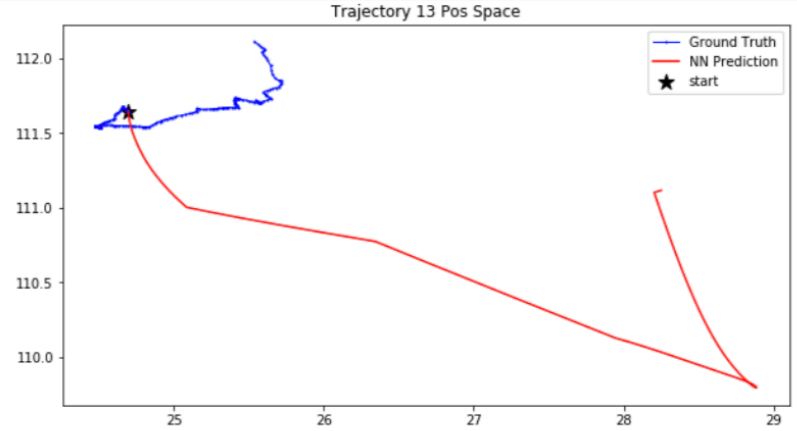

Trajectory 2 Pos Space
Ground Truth
NN Prediction
start

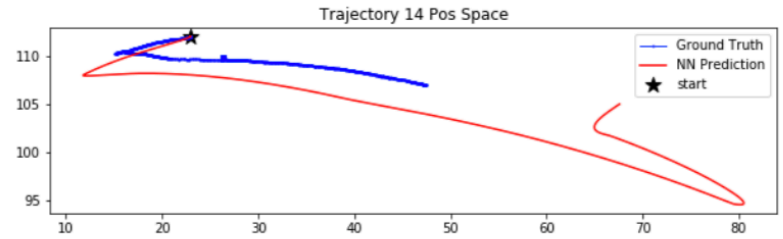

Trajectory 3 Pos Space
Ground Truth
NN Prediction
start

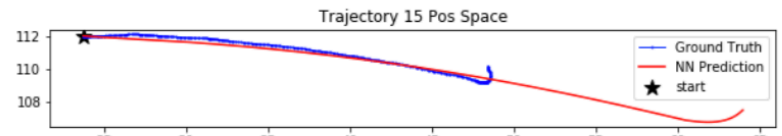

Trajectory 4 Pos Space
Ground Truth
NN Prediction
start

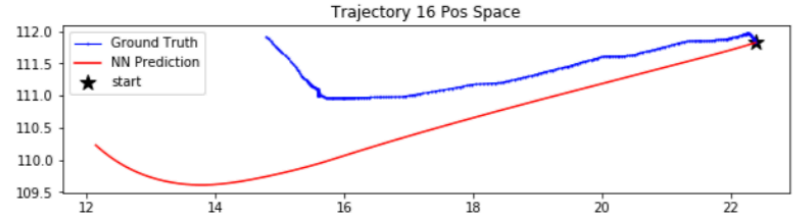

Trajectory 5 Pos Space
Ground Truth
NN Prediction
start

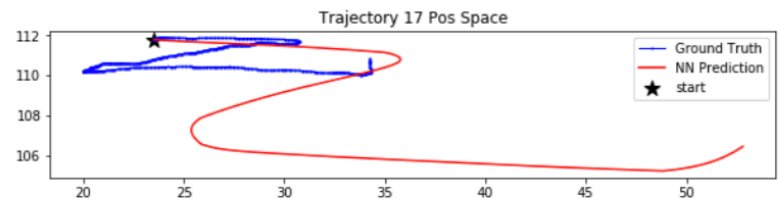

Trajectory 6 Pos Space
Ground Truth
NN Prediction
start

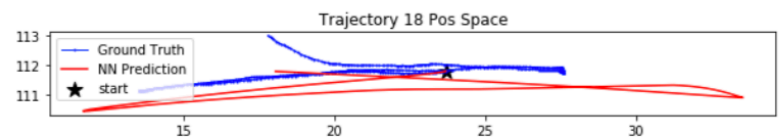

Trajectory 13 Pos Space
Ground Truth
NN Prediction
start


Trajectory 14 Pos Space
Ground Truth
NN Prediction
start

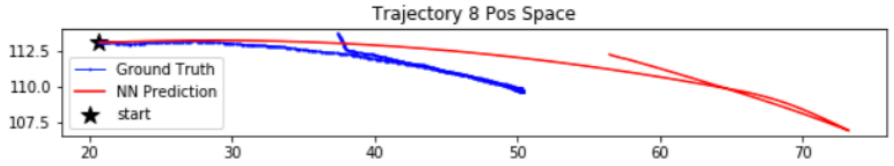

Trajectory 15 Pos Space
Ground Truth
NN Prediction
start

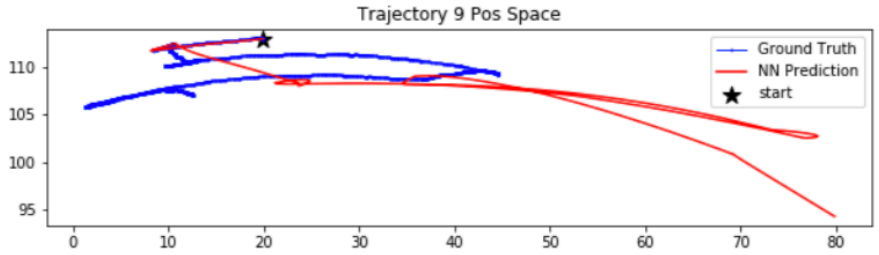

Trajectory 16 Pos Space
Ground Truth
NN Prediction
start

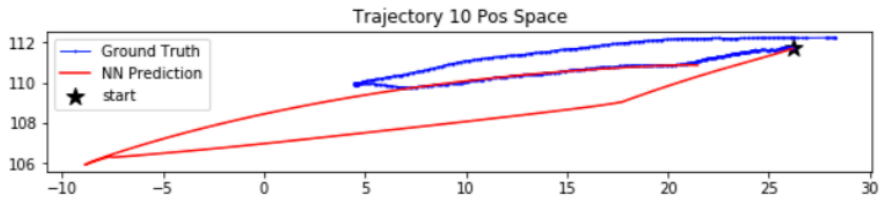

Trajectory 17 Pos Space
Ground Truth
NN Prediction
start

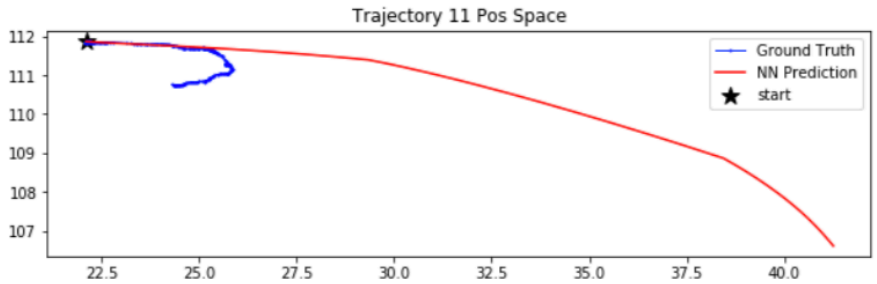

Trajectory 18 Pos Space
Ground Truth
NN Prediction
start

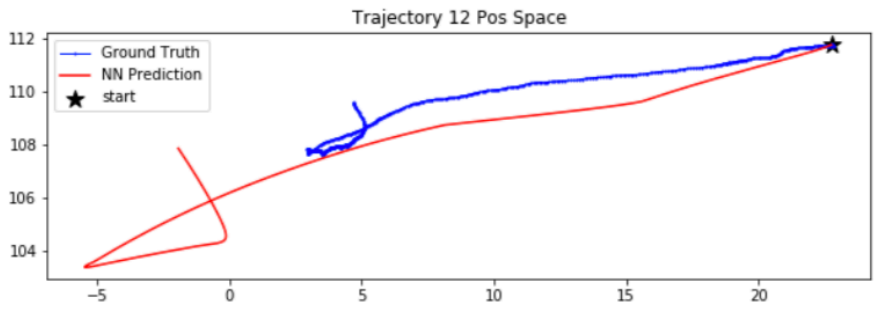

Trajectory 7 Pos Space
Ground Truth
NN Prediction
start

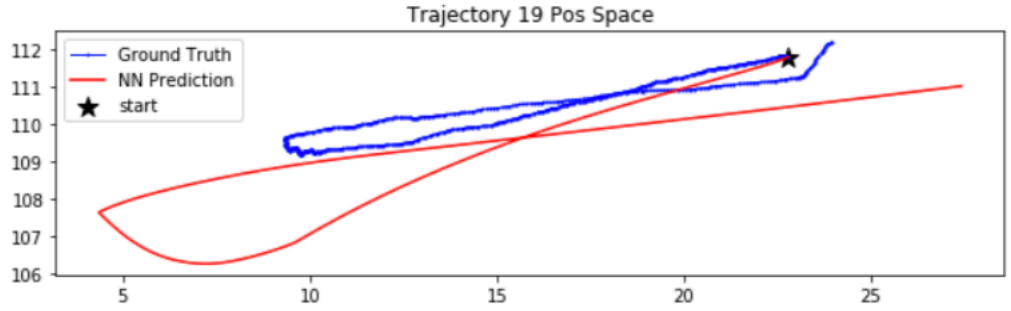

Trajectory 8 Pos Space
Ground Truth
NN Prediction
start

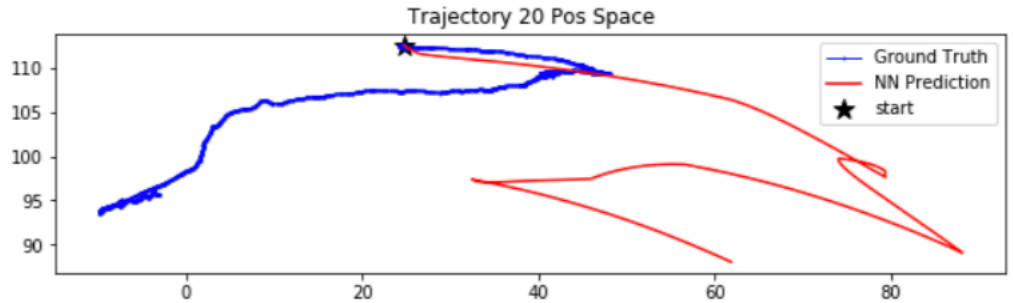

Trajectory 9 Pos Space
Ground Truth
NN Prediction
start

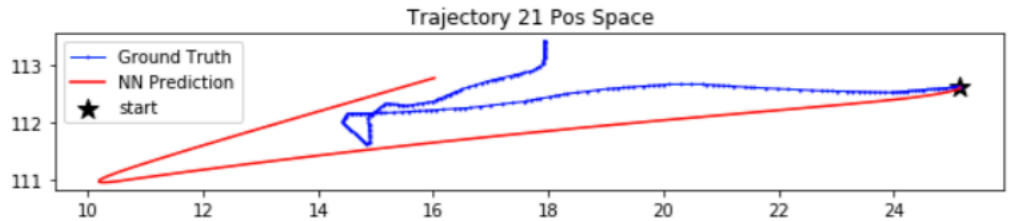

Trajectory 10 Pos Space
Ground Truth
NN Prediction
start


Trajectory 11 Pos Space
Ground Truth
NN Prediction
start

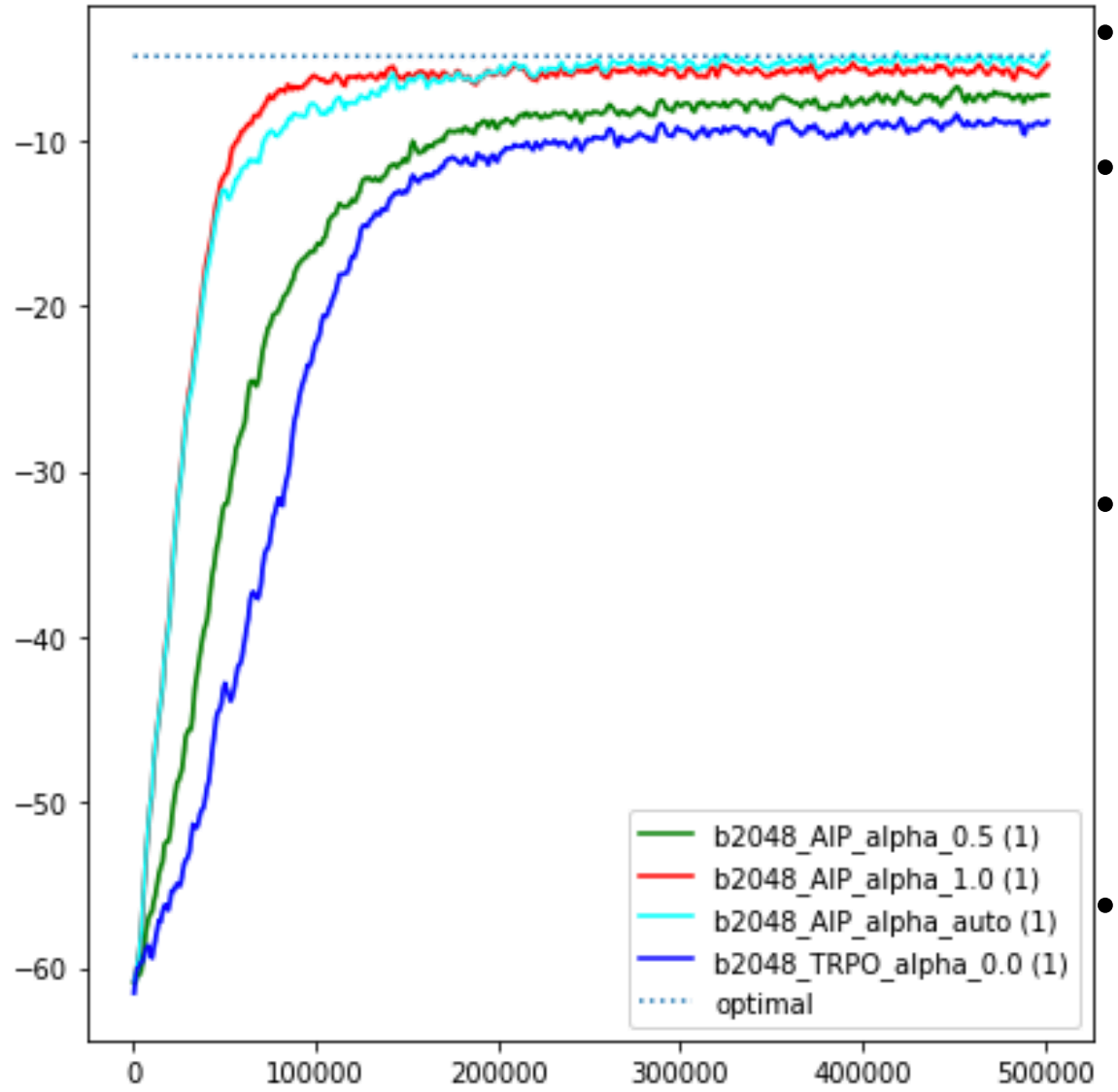

Trajectory 12 Pos Space
Ground Truth
NN Prediction
start


Trajectory 19 Pos Space
Ground Truth
NN Prediction
start

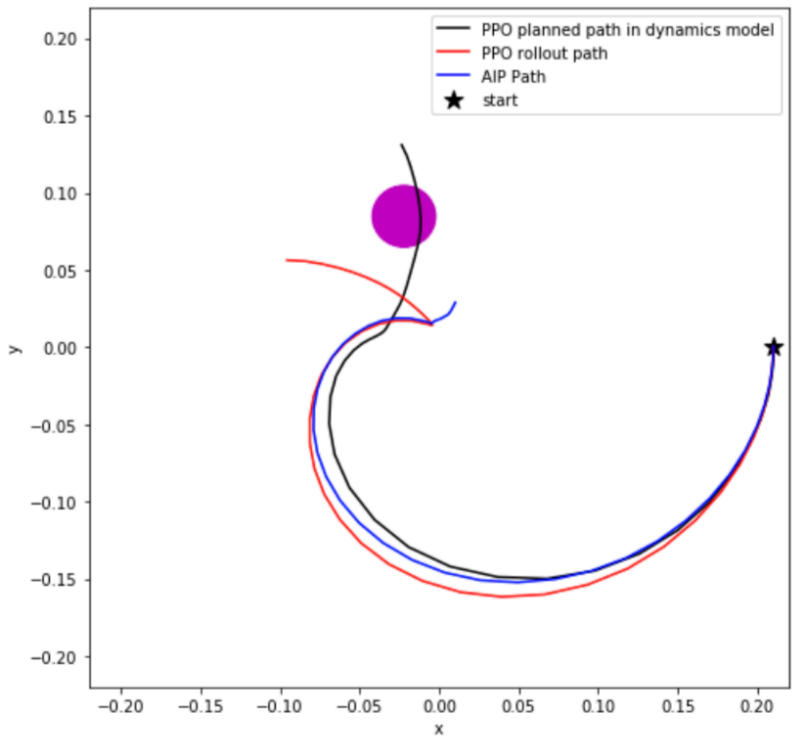

Trajectory 20 Pos Space
Ground Truth
NN Prediction
start

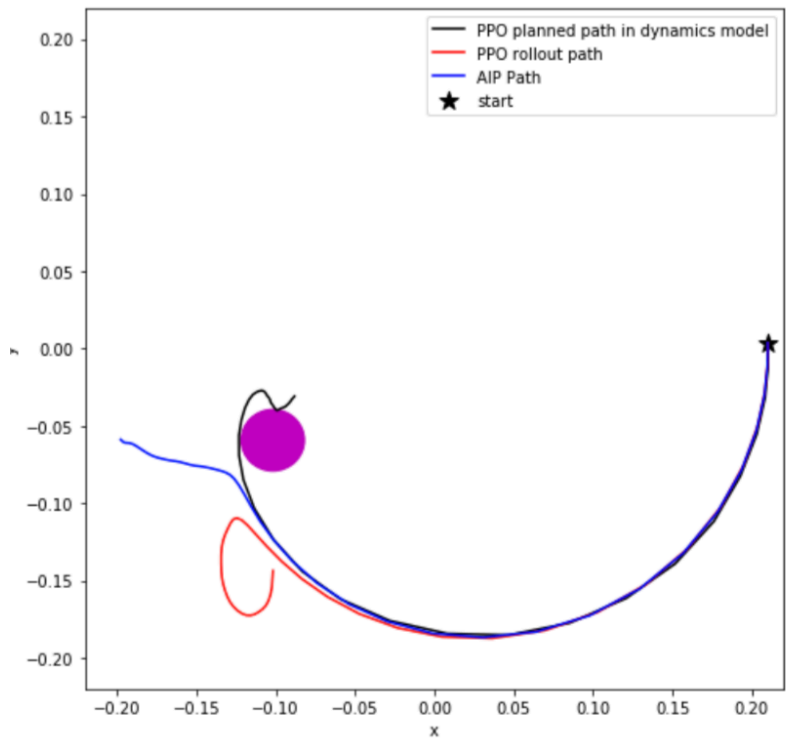

Trajectory 21 Pos Space
Ground Truth
NN Prediction
start

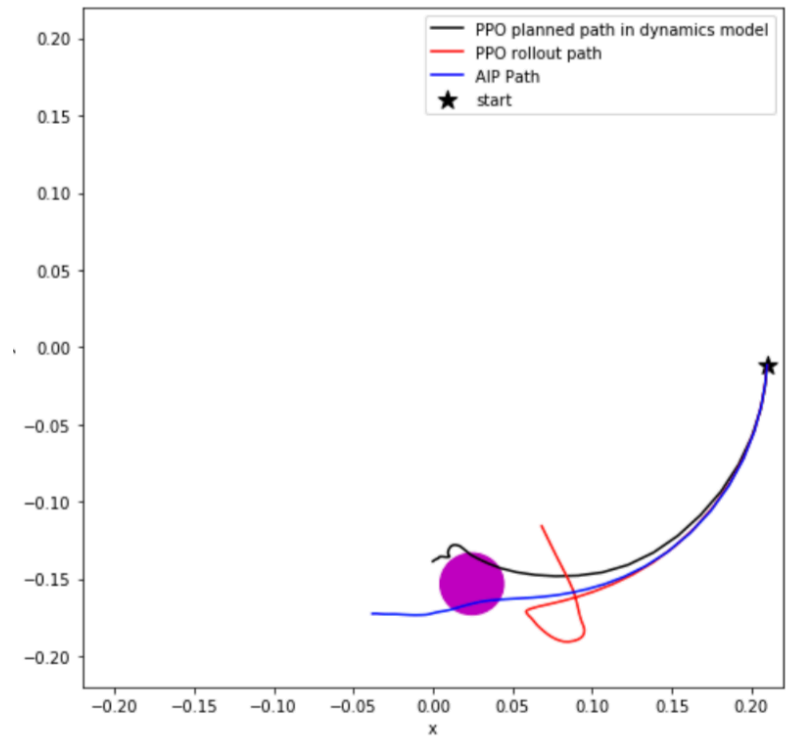
Mujoco-Reacher (Ablation Study)



- Compared to TRPO(alpha=0), AIP is far more sample-efficient.
- Compared to pure model-based method(alpha=1.0) which achieves a best return of -5.6, our AIP shows a better performance with an approximately optimal return of -4.9, which could also be achieved by TRPO, however, only after 1 million timesteps.
- AIP works better than pure model-based method(alpha=1.0), mainly because it does not throw away the exploration. Finally, AIP finds a better policy than pure model-based method(alpha=1.0). In the case of gazebo hand with obstacles, the difference between AIP and \alpha=1.0 could be more obvious.
- AIP could weigh the exploration(a_explore term in the Gaussian mean) and exploitation(a_ref term from model-based policy) intelligently using a weight \alpha dependent on the reward uncertainty r_diff.


PPO planned path in dynamics model
PPO rollout path
AIP Path
start
x
y


PPO planned path in dynamics model
PPO rollout path
AIP Path
start
x


PPO planned path in dynamics model
PPO rollout path
AIP Path
start
x

Gazebo Hand (Result is not yet available)

- Encountered many problems when implementing AIP on ros Gazebo
- Many library version dismatches (from python3 to python2, since gazebo ros is built on python2) (e.g. pickle, tensorflow, pytorch)
- Some ros issues when running ros, such as definition of ros message type
- Could be solved soon

Next Steps

- Focus on :

  Gazebo Hand: AIP (by this weekend)

  Real Hand: A* rollout, PPO rollout, LQR closed-loop
  (by this weekend/beginning next week)

- Remaining:

  Acrobot: - AIP of sparse reward -1, (adapt r_diff in terms of long-term Q=r+V(s'))

  - LQR of discrete actions

  Real hand: AIP (if we have enough time)