[6.1.2021]

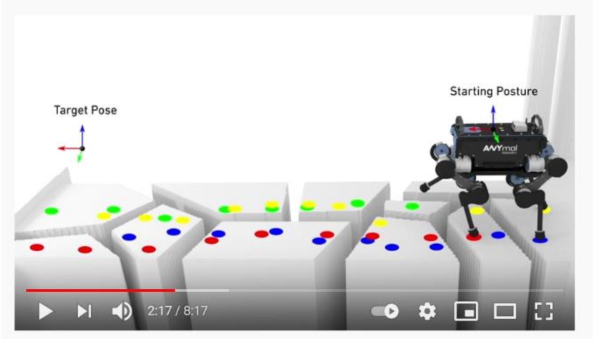
[DeepGait: Planning and Control of Quadrupedal Gaits Using Deep Reinforcement Learning]

Summary

Along with the paper there is a fascinating video, although it was still a simulation video but we expect the experiments results will come out too.

https://www.youtube.com/watch?v=jIKhnWzcdbg

- The problem setting: In such terrains, how to decide the footholds and how to generate the related base trajectories and feet trajectories (which is often neglected by me)



- One observation is that this robot is able to adjust to different gaps, one reason is that the phase state consists of two times t_E, t_S that can be used to predict the feed trajectory.
- The phase state (that is to be tracked by the gait controller) has information about foot position and base position, those are extracted for Gait controller.
- Planner plan multiple steps, controller just takes one step to track, GP updates desired foothold at around 2 Hz (I guess some time is spent on the LP transition feasibility criteria part), GC computes the foothold tracking error (as part of

observation) at around 100 Hz, the PD controller that tracks the joints given by GC is about 400 Hz.

- Question? What if we reduce the length of planning?

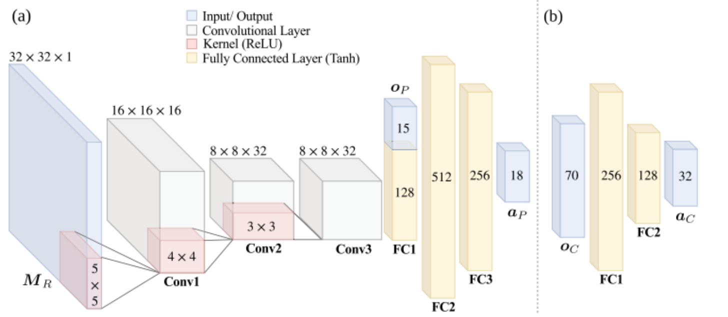
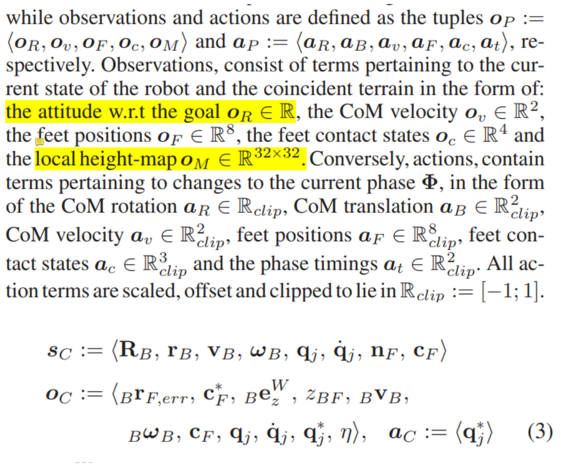


Fig. 3. The neural-network models used for the latent parameters of policy distributions of the (a) GP and (b) GC, respectively.



while observations and actions are defined as the tuples $\boldsymbol{o}_P := \langle \boldsymbol{o}_R, \boldsymbol{o}_v, \boldsymbol{o}_F, \boldsymbol{o}_c, \boldsymbol{o}_M \rangle$ and $\boldsymbol{a}_P := \langle \boldsymbol{a}_R, \boldsymbol{a}_B, \boldsymbol{a}_v, \boldsymbol{a}_F, \boldsymbol{a}_c, \boldsymbol{a}_t \rangle$, respectively. Observations, consist of terms pertaining to the current state of the robot and the coincident terrain in the form of: the attitude w.r.t the goal $\boldsymbol{o}_R \in \mathbb{R}$, the CoM velocity $\boldsymbol{o}_v \in \mathbb{R}^2$, the feet positions $\boldsymbol{o}_F \in \mathbb{R}^8$, the feet contact states $\boldsymbol{o}_c \in \mathbb{R}^4$ and the local height-map $\boldsymbol{o}_M \in \mathbb{R}^{32\times 32}$. Conversely, actions, contain terms pertaining to changes to the current phase $\boldsymbol{\Phi}$, in the form of the CoM rotation $\boldsymbol{a}_R \in \mathbb{R}_{clip}$, CoM translation $\boldsymbol{a}_B \in \mathbb{R}^2_{clip}$, CoM velocity $\boldsymbol{a}_v \in \mathbb{R}^2_{clip}$, feet positions $\boldsymbol{a}_F \in \mathbb{R}^8_{clip}$, feet contact states $\boldsymbol{a}_c \in \mathbb{R}^3_{clip}$ and the phase timings $\boldsymbol{a}_t \in \mathbb{R}^2_{clip}$. All action terms are scaled, offset and clipped to lie in $\mathbb{R}_{clip} := [-1; 1]$.

$$
\begin{aligned}
\boldsymbol{s}_C &:= \langle \mathbf{R}_B,\ \mathbf{r}_B,\ \mathbf{v}_B,\ \boldsymbol{\omega}_B,\ \mathbf{q}_j,\ \dot{\mathbf{q}}_j,\ \mathbf{n}_F,\ \mathbf{c}_F \rangle \\
\boldsymbol{o}_C &:= \langle {}_B\mathbf{r}_{F,err},\ \mathbf{c}^*_F,\ {}_B\mathbf{e}^W_z,\ z_{BF},\ {}_B\mathbf{v}_B, \\
&\qquad {}_B\boldsymbol{\omega}_B,\ \mathbf{c}_F,\ \mathbf{q}_j,\ \dot{\mathbf{q}}_j,\ \mathbf{q}^*_j,\ \eta \rangle, \quad \boldsymbol{a}_C := \langle \mathbf{q}^*_j \rangle \qquad (3)
\end{aligned}
$$

  - Why 70 and 32?
    - 70 = 3+4+3+1+3+3+4+12+12+24+1
    - 32 = 24+? Or just some part of the output is not used?

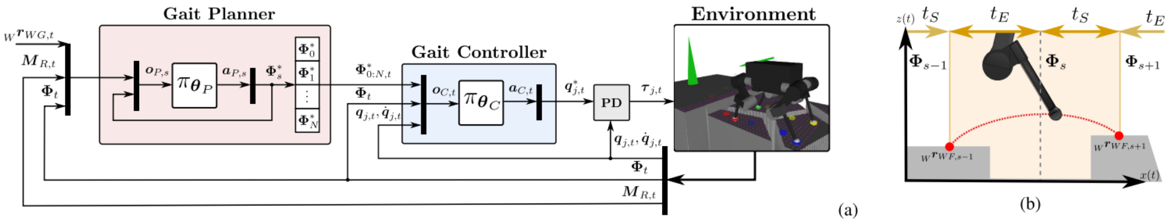


Fig. 2. (a) Overview of the proposed control structure used at deployment time. (b) Phases within a sequence are indexed using s, and every index corresponds to a point in time centered around a window defined by the durations t_E and t_S. The center of the window is defined by the motion of the base as captured by the phase $\boldsymbol{\Phi}_s$. t_S defines the time-to-switch from the current contact support to the next, specified in $\boldsymbol{\Phi}_{s+1}$, and t_E defines the time elapsed since the switch from the previous contact support, specified in $\boldsymbol{\Phi}_{s-1}$, to the current.

Thoughts

- This paper assumes full information about the terrain → a big assumption so far, maybe in the future work this part will also be learned.
  - The what's the difference between this paper with the science paper that use teacher-student transfer learning where real data is also collected and the higher-level control is also simply the moving direction command?

- How the map information is incorporated into the network?
    - It is added into the input of the policy → the observation of the policy is generated by concatenating the latent representation of map with raw proprioceptive measurements.
- Contact-phase → transition feasibility criteria, instead of using physical interactions. →check another paper *C-CROC: Continuous and convex resolution of centroidal dynamic trajectories for legged robots in multi-contact scenarios*
    - Transition feasibility criteria aims at determining if there exists a valid solution that could connecting two states → using parametrization for center position to solve the feasibility problem by a LP problem.
    - *The intermediate kinematic constraints are evaluated by interpolating (not important part)*
- The policy infers the distribution of the phase function, how to choose one? Softmax?
    - This is a general RL question, RL will based on the probability randomly choose the action and roll out the episode.
- Why we need deep reinforcement learning here? Network wise, what advantage does the deepness bring?
    - One reason mentioned in the paper is that because we use terrain map as input/observation here (local map, 32x32), so a large network is used.
    - Another reason is also because the output of the network policy (action space) here is large in dimension: 18
    - This question is only about the gait planner policy.
- How is the dynamics property being depicted? Looks like the robot is always in a semi-static status.
    - The planning part is still kinodynamic, because we only consider the phase feasibility, and it is discrete.
    - The controller part tracks the foothold, and it has to be dynamics stable, the physics in incorporated in Gait controller in term of transition dynamic function $T(s_{t-1}, s_t, a_t)$