PRELUDE Learning Poster

Tags: Perceptive Locomotion Planning, Robot Learning, Imitation Learning

(Reference: Seo, Mingyo, et al. “Learning to Walk by Steering: Perceptive Quadrupedal Locomotion in Dynamic Environments.” arXiv preprint arXiv:2209.09233 (2022).)

Problem: Perceptive Quadrupedal Locomotion Planning **Training Dataset:**

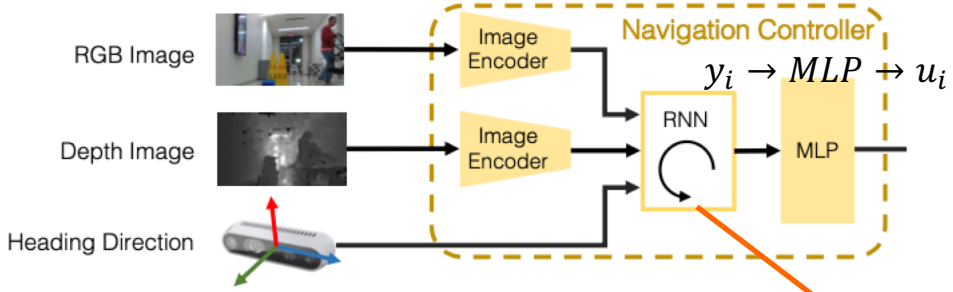
Solution:

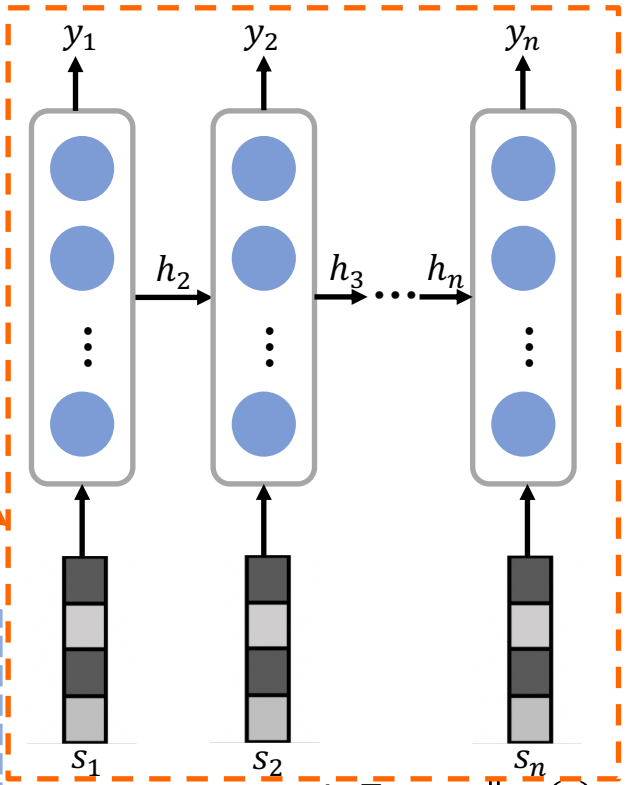
High-level imitation learning:

Human demonstration: $\mathcal{D} = \{s_i, u_i\}_{i=1}^{N}$

Low-level RF learning: Simulated data

1. Model problem as discrete-time MDP problem:
   $\mathcal{M}(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \rho_0)$
   learn policy π: $a_t = \pi(s_t)$ to minimize $\mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t, s_{t+1})]$
2. Hierarchical policy learning framework: $a_t = \pi(s_t) = \pi_L(s_t, \pi_H(s_t))$
   a. High-level decision-making to predict navigation commands:
   $u_t = \pi_H(s_t) \rightarrow$ Imitation Learning:



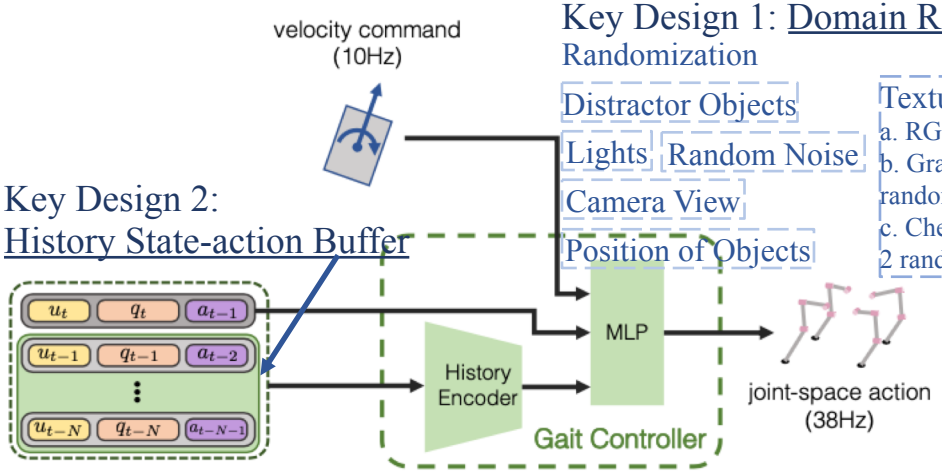


$\arg\min_{\theta} \mathbb{E}_{(s,a)\sim\mathcal{D}} \|\pi_\theta(s) - a\|^2$

Behavioral Cloning Model with RNNs:

For capturing temporal info of moving objects (eg. moving people).

   b. Low-level gait generation:
   $a_t = \pi_L(s_t, u_t) \rightarrow$ Reinforcement Learning:



Summarization and Personal Thinking

Summarization of innovations:

1. Hierarchical policy learning in MDP problem
2. For high-level imitation learning:
   a. Utilize behavioral cloning model with RNNs to capture temporal info of moving objects.
3. For low-level reinforcement learning:
   a. Domain randomization: Provide enough variability in simulator to bridge ‘reality gap’.

   Reference: Tobin, Josh, et al. "Domain randomization for transferring deep neural networks from simulation to the real world."

   b. Historical state-action buffer: Robot’s recent state-action history can serve as a robust proxy for estimating.

Questions:

1. Since limitations on the amount of human demonstration data, lack of more general human actions in different terrains and complexity of model compared with training data amount, can high-level imitation learning generalize to more general and sophisticated navigation scenes?
2. Can eliminate MLP layer in imitation learning model and totally consider this as time-sequential problem with temporal inputs? (Thus, models like RNN or Transformer can fit for total problem)

Personal thinking of future work:

1. Effective data augmentation in human demonstration dataset for imitation learning.
2. To adapt the model complexity to the magnitude of the training data to avoid overfitting and obtain better generalization, the imitation learning model needs to be pruned properly.
3. Introduce expert intervention in low-level RF learning to help improve effectiveness of gait generation as well as add truly realistic variability at same time.

Reference: Spencer, et al. Expert Intervention Learning: An online framework for robot learning from explicit and implicit human feedback.