Skill Tuning in Pretrained Skill-Conditioned Policies

Rob Harries

CS 224R
Deep RL

Environment

Mujoco Walker_Walk
Two-Dimensional Locomotion

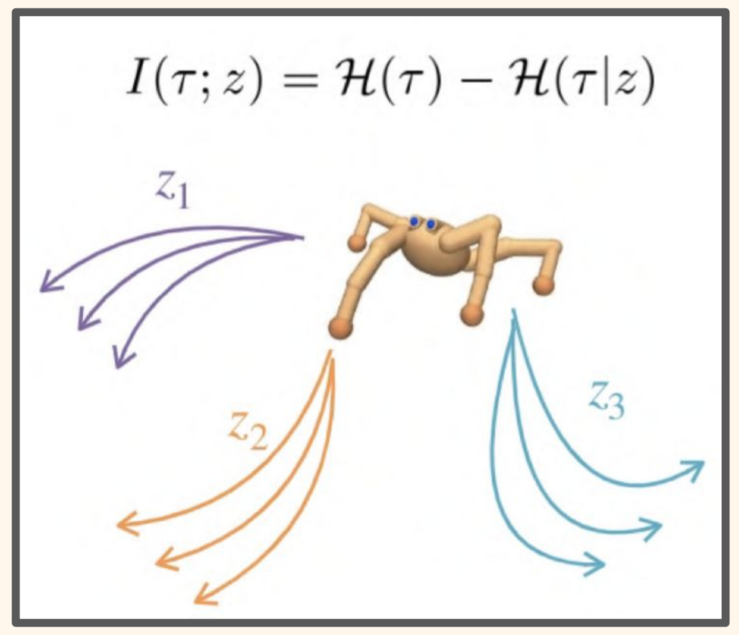
Rewarded for:

- Specific Forward Velocity
- Upright and Elevated Torso

$$r = \left(\frac{3}{4} r_{\text{torso\_elevated}} + \frac{1}{4} r_{\text{torso\_upright}}\right)\left(\frac{1}{6} + \frac{5}{6} r_{\text{forward\_vel}}\right)$$

Reward-Free Pretraining

Contrastive Intrinsic Control (2022)



Maps randomized skill vectors to diverse, predictable trajectories

Maximizes Mutual Info between skill, trajectory

Maximizes dissimilarity between different skills

$$z \in \mathbb{R}^{64}$$

2,000,000 training steps

High-dimensional continuous skill space

Task-Specific Skill Initialization

- Use first 4,000 steps in task to test return on 40 different skills spread along diagonal:
  z = <0, 0, ..., 0> to z = <1, 1, ..., 1>
- Default method selects and freezes the best performing skill, then finetunes the actor
- In Skill Vocab Tuning, top K skills are chosen as initialization, then further trained

Can pretrained policies solve tasks without finetuning?

<u>Default:</u> Select a fixed skill vector and finetune skill-conditioned policy.

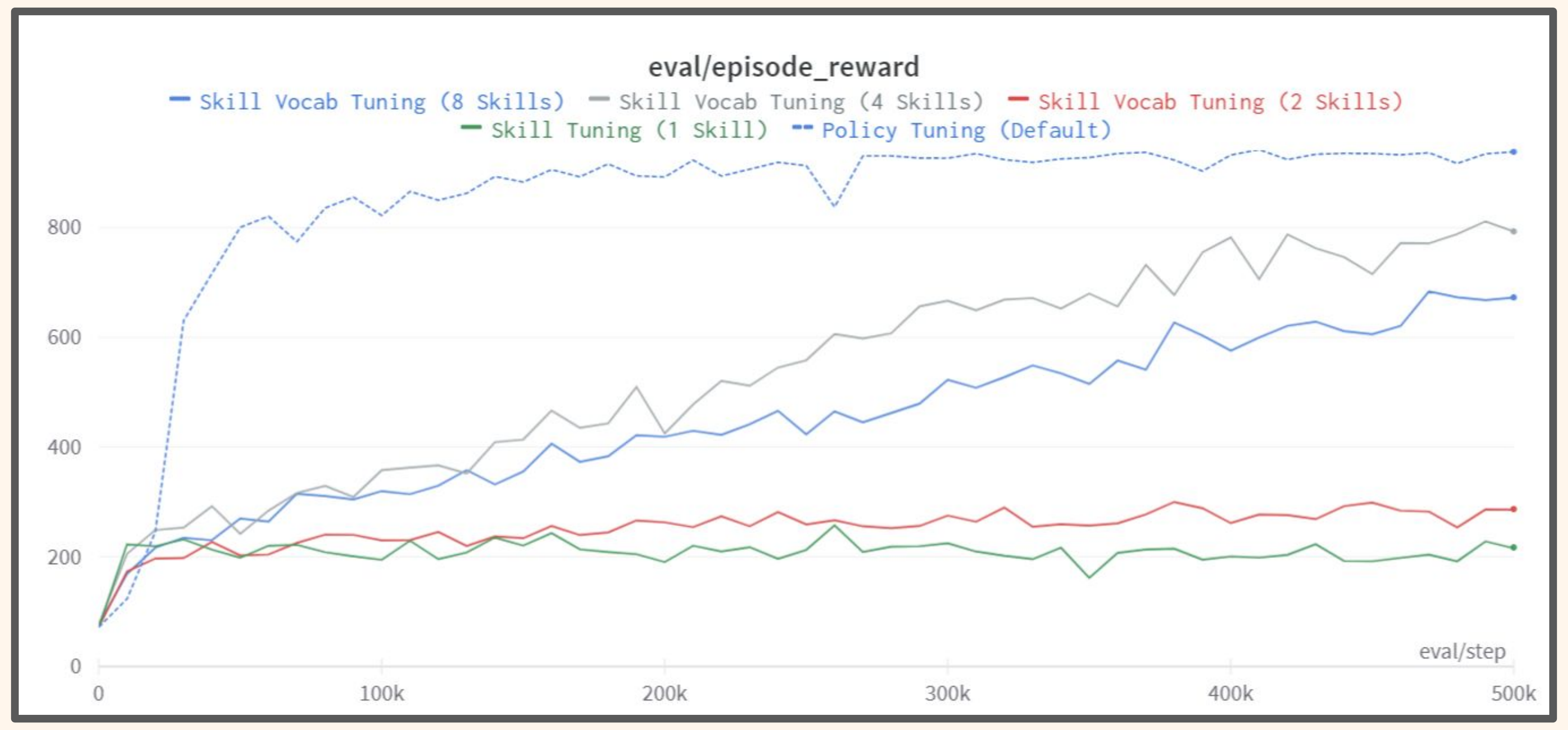
<u>Option A</u>: Freeze policy, tune input skill vector.

<u>Option B:</u> Freeze policy, tune multiple skill vectors alongside a skill-selection policy.

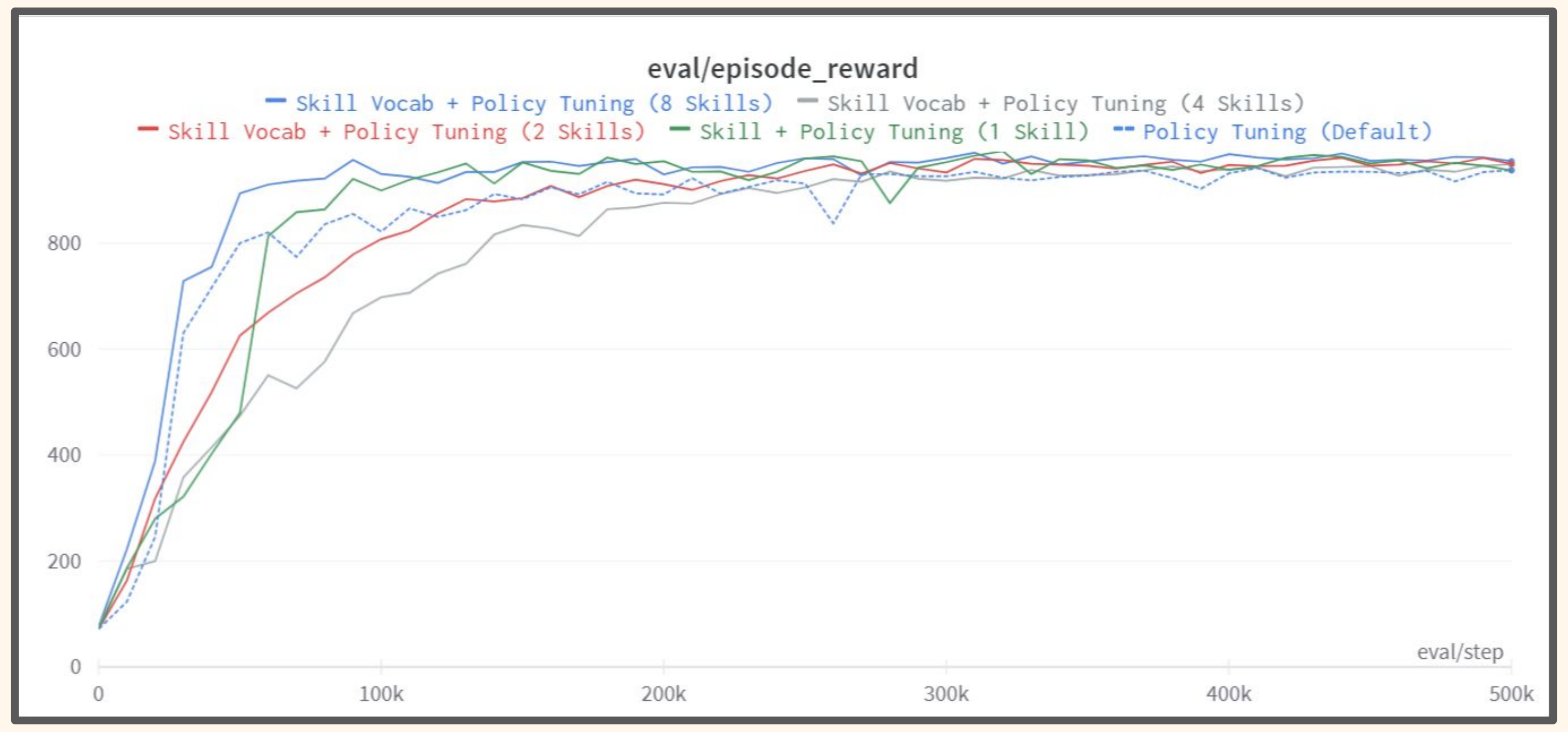
Results

<u>Option A</u>: Single Skill Tuning cannot solve task.

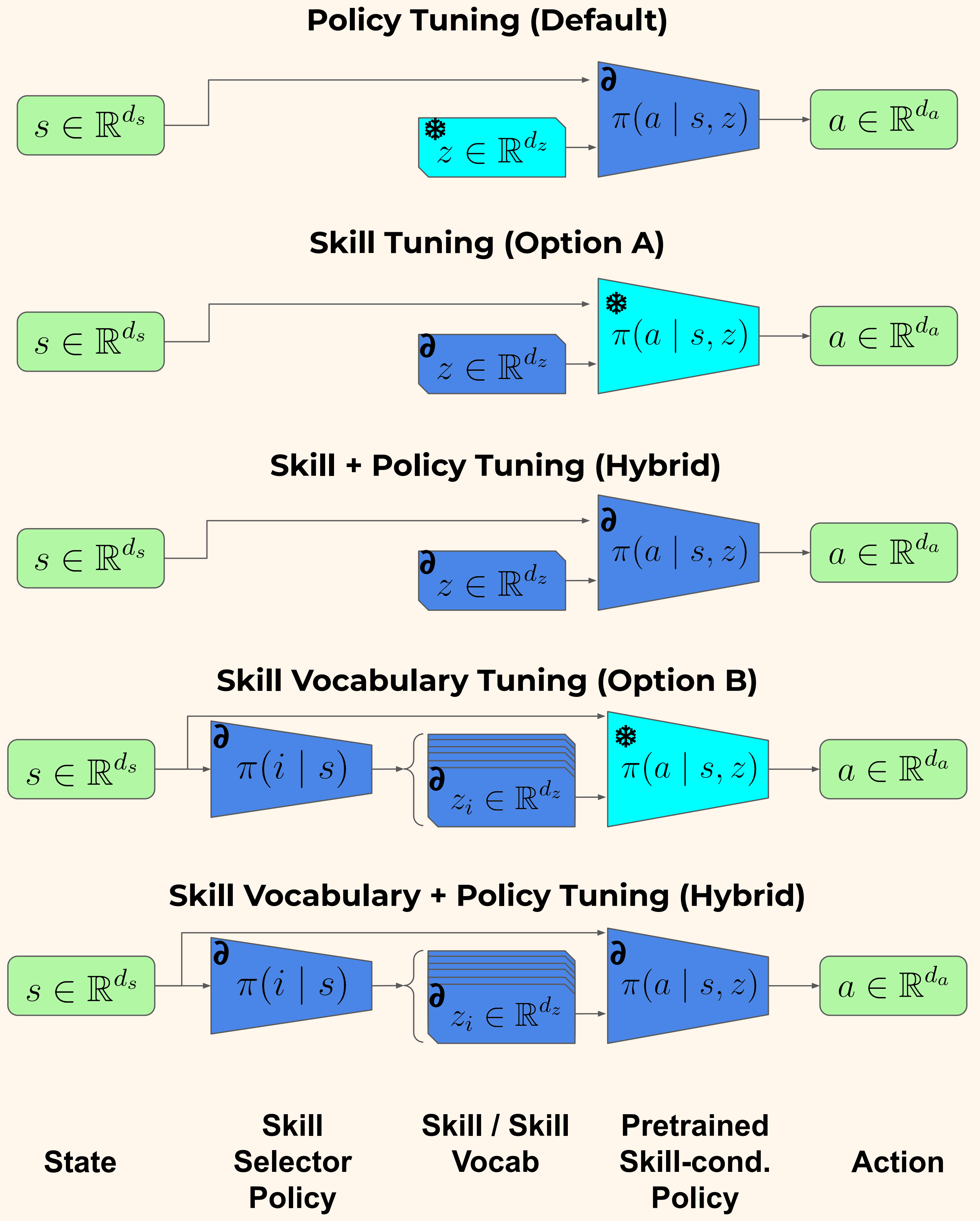
<u>Option B:</u> Skill Vocabulary Tuning can solve task, albeit inefficiently.



<u>Hybrid:</u> Simultaneous skill + policy tuning yields slight efficiency improvements (using 8 skills).



Methods



Policy Tuning (Default)

Skill Tuning (Option A)

Skill + Policy Tuning (Hybrid)

Skill Vocabulary Tuning (Option B)

Skill Vocabulary + Policy Tuning (Hybrid)

State

Skill Selector Policy

Skill / Skill Vocab

Pretrained Skill-cond. Policy

Action

DDPG Actor Loss Function

Single Skill

$$\mathcal{L}_{\text{actor}}(\theta, z) = \mathbb{E}_s\left[-Q^{\phi}\left(s, \pi^{\theta}(s, z)\right)\right]$$

Multiple Skills

$$\mathcal{L}_{\text{actor}}(\theta, z, \gamma) = \mathbb{E}_s\left[-\sum_{i=1}^{n} \pi^{\gamma}(i \mid s) Q^{\phi}(s, \pi^{\theta}(s, z_i))\right]$$
$$+\lambda \mathbb{E}_s\left[\mathcal{H}(\pi^{\gamma}(\cdot \mid s))\right]$$