EXTRACT: Efficient Policy Learning by Extracting Transferrable Robot Skills

Jesse Zhang[1], Minho Heo[2], Zuxin Liu[3],
Erdem Bıyık[1], Joseph J. Lim[2], Yao Liu[4], Rasool Fakoor[4]
[1]University of Southern California, [2]KAIST, [3]CMU, [4]Amazon Web Services

Abstract

Reinforcement learning (RL) agents equipped with useful, temporally extended skills *can learn new tasks more easily. Prior work in skill-based RL either requires expert supervision to define useful skills or creates non-semantically aligned skills from offline data through heuristics, which is difficult for a downstream RL agent to use for learning new tasks. Instead, our approach, EXTRACT, utilizes pretrained vision models to extract a discrete set of semantically meaningful skills from offline data, each of which is parameterized by continuous arguments,* without human supervision*. This skill parameterization allows robots to learn new tasks more quickly by only needing to learn* when *to select a specific skill and* how *to modify its arguments for the specific task. We demonstrate through experiments in sparse-reward, image-based, robot manipulation environments, both in simulation and in the real world, that EXTRACT can more quickly learn new tasks than prior skill-based RL, with up to a* $\mathbf{10}\times$ *gain in sample efficiency.*

1. Introduction

Imagine learning to play racquetball as a complete novice. Without prior experience in racket sports, this poses a daunting task that requires learning not only the (1) complex, high-level strategies to control *when* to serve, smash, and return the ball but also (2) *how* to actualize these moves in terms of fine-grained motor control. However, a squash player should have a considerably easier time adjusting to racquetball as they already know how to serve, take shots, and return; they simply need to learn *when* to use these skills and *how* to adjust them for larger racquetball balls. In this paper, we aim to utilize this intuition to enable efficient learning of new tasks.

In general, humans can learn new tasks quickly—given prior experience and mastery of relevant skills—by adjusting existing skills for the new task [2, 4]. Skill-based reinforcement learning (RL) aims to emulate this efficient transfer learning [1, 3, 8, 12, 13, 15, 18–20] in learned agents by equipping them with a wide range of skills (i.e., temporally-extended action sequences) that they can call upon for efficient downstream learning. Using skills instead of unstructured, low-level actions, skill-based RL reduces task time horizons and yields more effective exploration. However, existing skill-based RL approaches rely on costly human supervision [3, 9, 12, 16] or restrictive definitions of skills [1, 6, 13] that limit the expressiveness and adaptability of the learned skills. Therefore, we ask: how can robots discover *adaptable* skills for efficient transfer learning *without costly human supervision*?

Calling back to the squash to racquetball transfer example, we humans categorize different racket movements into *discrete skills*—for example, a "forehand swing" is distinct from a "backhand return." These discrete skills can be directly transferred by making minor modifications for racquetball's larger balls and different rackets. This process is akin to that of calling a programmatic API, e.g., `def forehand(x, y)`, where learning to transfer reduces to learning *when* to call discrete functions (e.g., `forehand()` vs `backhand()`) and *how* to execute them (i.e., what their arguments should be). In this paper, we propose a method to accelerate transfer learning by enabling robots to learn, without expert supervision, a discrete set of skills parameterized by input arguments that are useful for downstream tasks. We assume access to a general offline dataset containing image-action pairs trajectories but not the downstream target tasks. Our key insight is aligning skills by extracting *high-level behaviors* from trajectory images, i.e., discrete skills like "forehand swing," contained within the dataset. Specifically, we use video encoders from pretrained vision-language models (VLMs), which are trained to align images with language descriptions [14] so that images of similar high-level behaviors are embedded to similar latent embeddings [17]. However, two challenges preclude realizing this insight: (1) how to align individual embeddings into a set of discrete, input-parameterized skills, and (2) how to guide online learning of new tasks with these skills.

To this end, we propose EXTRACT (**Ex**traction of **T**ransferrable **R**obot **Act**ion Skills), a framework for extracting discrete, parameterized skills from offline data to

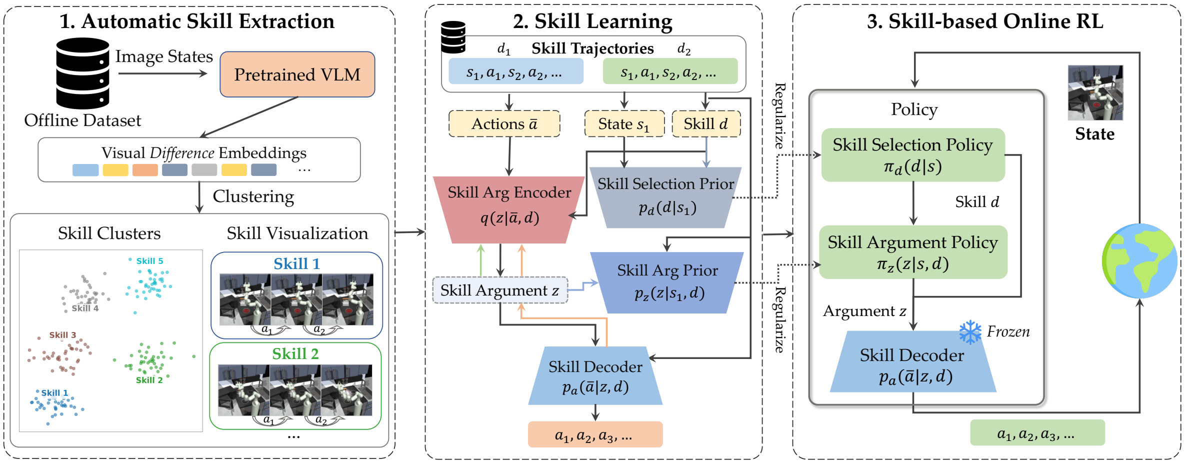
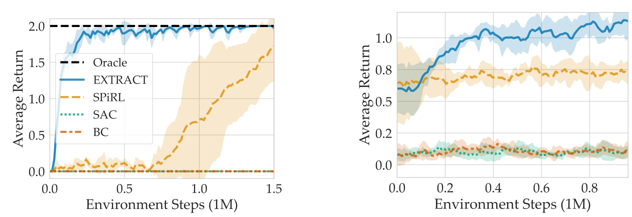


Figure 1. EXTRACT consists of three phases to enable efficient transfer learning. **(1) Skill Extraction**: We extract a set of high-level skills from offline robot interaction data by clustering together visual difference embeddings, representing changes in high-level behaviors of images in each trajectory; here, each cluster corresponds to a high-level behavior (skill). **(2) Skill Learning**: We aim to obtain a skill decoder model, $p_a(\bar{a} \mid z, d)$, to output variable-length action sequences conditioned on a skill ID d and a learned continuous argument z. The argument z is learned by training $p_a(\bar{a} \mid z, d)$ with a VAE reconstruction objective from action sequences encoded by a skill encoder, $q(z \mid \bar{a}, d)$, conditioned on the action sequence and skill ID d. We additionally train a skill selection prior and skill argument prior $p_d(d \mid s)$, $p_z(z \mid s, d)$ to predict which skills d and their arguments z are useful for a given state s. Colorful arrows indicate gradients from reconstruction, argument prior, selection prior, and VAE regularization losses. **(3) Online RL**: To learn a new task, we train a skill selection and skill argument policy with RL while regularizing them with the skill selection and skill argument priors. These skills and arguments are given to the skill decoder, $p_a(\bar{a} \mid z, d)$, and translated into low-level actions to be executed in the environment.

guide online learning of new tasks (see Figure 1). We first use a pre-trained VLM to extract observation embedding differences, representing changes in high-level behaviors over time (i.e., $VLM(s_t) - VLM(s_1)$), of offline trajectories. Next, we cluster the difference embeddings in an *unsupervised* manner to form discrete skill clusters that represent high-level skills. To parameterize these skills, we train a *skill decoder* on these clusters, conditioned on the skill ID (e.g., representing a "backhand return") and a learned argument (e.g., indicating position and velocity), to produce a skill consisting of a temporally extended, variable-length action sequence. Finally, to train a robot for new tasks, we train a skill-based RL policy to act over this skill-space while being guided by skill prior networks, learned from our offline skill data, guiding the policy for (1) *when* to select skills and (2) *what* their arguments should be.

2. Experiments

We evaluate EXTRACT on **Franka Kitchen** [5] and **LIBERO-10** [10], two long-horizon, image-based, robot manipulation benchmarks with sparse rewards. For both, we pre-train on scripted or human teleoperation trajectories and evaluate on unseen, long-horizon tasks. We compare against: (1) an **Oracle** [3] which is given discrete, human-designed skills; (2) **SPiRL** [13], which randomly segments sequences of actions into a continuous skill-space; (3) **BC**, behavior cloning; and (4) **SAC** [7] as the standard RL baseline.



(a) Franka Kitchen (b) LIBERO-10

Figure 2. Online RL results.

Our method **EXTRACT** uses the R3M VLM [11] and K-means clustering with $K = 8$ for offline skill extraction. Finally, all reported experimental results are means and standard deviations over 3 seeds.

Results. We can see in Figure 2 that EXTRACT is **10x** more sample-efficient than SPiRL, in yellow, and matches the Oracle skill (RAPS [3]) method performance in Franka Kitchen. In LIBERO-10, EXTRACT also outperforms all other methods, achieving **2x** the final performance of SPiRL. This improvement of our method over SPiRL is likely due to two reasons: on average, longer skills and a semantically structured discrete skill space instead of the random latent skills that SPiRL learns.

References

[1] Anurag Ajay, Aviral Kumar, Pulkit Agrawal, Sergey Levine, and Ofir Nachum. {OPAL}: Offline primitive discovery for accelerating offline reinforcement learning. In *International Conference on Learning Representations*, 2021. 1

[2] John R. Anderson. Acquisition of cognitive skill. *Psychological Review*, 89:369–406, 1982. 1

[3] Murtaza Dalal, Deepak Pathak, and Ruslan Salakhutdinov. Accelerating robotic reinforcement learning via parameterized action primitives. In *NeurIPS*, 2021. 1, 2

[4] P.M. Fitts and M.I. Posner. *Human Performance*. Brooks/-Cole Publishing Company, 1967. 1

[5] Justin Fu, Aviral Kumar, Ofir Nachum, George Tucker, and Sergey Levine. D4rl: Datasets for deep data-driven reinforcement learning. *arXiv preprint arXiv:2004.07219*, 2020. 2

[6] Abhishek Gupta, Vikash Kumar, Corey Lynch, Sergey Levine, and Karol Hausman. Relay policy learning: Solving long-horizon tasks via imitation and reinforcement learning. *CoRL*, 2019. 1

[7] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *ICML*, 2018. 2

[8] Karol Hausman, Jost Tobias Springenberg, Ziyu Wang, Nicolas Heess, and Martin Riedmiller. Learning an embedding space for transferable robot skills. In *ICLR*, 2018. 1

[9] Youngwoon Lee, Shao-Hua Sun, Sriram Somasundaram, Edward S Hu, and Joseph J Lim. Composing complex skills by learning transition policies. In *International Conference on Learning Representations*, 2018. 1

[10] Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning, 2023. 2

[11] Suraj Nair, Aravind Rajeswaran, Vikash Kumar, Chelsea Finn, and Abhinav Gupta. R3m: A universal visual representation for robot manipulation. In *CoRL*, 2022. 2

[12] Soroush Nasiriany, Huihan Liu, and Yuke Zhu. Augmenting reinforcement learning with behavior primitives for diverse manipulation tasks. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2022. 1

[13] Karl Pertsch, Youngwoon Lee, and Joseph J. Lim. Accelerating reinforcement learning with learned skill priors. In *Conference on Robot Learning (CoRL)*, 2020. 1, 2

[14] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021. 1

[15] Stefan Schaal. Dynamic movement primitives–a framework for motor control in humans and humanoid robotics. *Adaptive Motion of Animals and Machines*, 2006. 1

[16] Kyriacos Shiarlis, Markus Wulfmeier, Sasha Salter, Shimon Whiteson, and Ingmar Posner. Taco: Learning task decomposition via temporal alignment for control. *ICML*, 2018. 1

[17] Sumedh Anand Sontakke, Jesse Zhang, Séb Arnold, Karl Pertsch, Erdem Biyik, Dorsa Sadigh, Chelsea Finn, and Laurent Itti. RoboCLIP: One demonstration is enough to learn robot policies. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 1

[18] Richard S. Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1):181–211, 1999. 1

[19] Jesse Zhang, Karl Pertsch, Jiefan Yang, and Joseph J Lim. Minimum description length skills for accelerated reinforcement learning. In *Self-Supervision for Reinforcement Learning Workshop - ICLR 2021*, 2021.

[20] Jesse Zhang, Karl Pertsch, Jiahui Zhang, and Joseph J. Lim. Sprint: Scalable policy pre-training via language instruction relabeling. In *International Conference on Robotics and Automation*, 2024. 1