Cognizing and Imitating Robotic Skills via a Dual Cognition-Action Architecture

Extended Abstract

Zixuan Chen
State Key Laboratory for Novel Software Technology
Nanjing University
Nanjing, China
chenzx@nju.edu.cn

Ze Ji
Cardiff University
Cardiff, United Kingdom
jiz1@cardiff.ac.uk

Shuyang Liu
State Key Laboratory for Novel Software Technology
Nanjing University
Nanjing, China
MG20330036@smail.nju.edu.cn

Jing Huo
State Key Laboratory for Novel Software Technology
Nanjing University
Nanjing, China
huojing@nju.edu.cn

Yiyu Chen
State Key Laboratory for Novel Software Technology
Nanjing University
Nanjing, China
yiyuiii@foxmail.com

Yang Gao
State Key Laboratory for Novel Software Technology
Nanjing University
Nanjing, China
gaoy@nju.edu.cn

ABSTRACT

Enabling robots to effectively learn and imitate expert skills in long-horizon tasks remains challenging. Hierarchical imitation learning (HIL) approaches have made strides but often fall short in complex scenarios due to their reliance on self-exploration. This paper introduces a novel approach inspired by the human skill acquisition process, proposing a Cognition-Action-based Robotic Skill Imitation Learning (CasIL) framework. CasIL integrates human cognitive priors for task decomposition into a dual-layer architecture, enhancing robots' ability to cognize and imitate essential skills from expert demonstrations. Our experiments across four RLbench tasks demonstrate CasIL's superior performance, robustness, and generalizability in skill imitation compared to related methods.

KEYWORDS

Hierarchical imitation learning, Robotic skill imitation, Visual demonstrations

ACM Reference Format:

Zixuan Chen, Ze Ji, Shuyang Liu, Jing Huo, Yiyu Chen, and Yang Gao. 2024. Cognizing and Imitating Robotic Skills via a Dual Cognition-Action Architecture: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

To advance robot skill imitation in long-horizon tasks, Hierarchical Imitation Learning (HIL) is recognized for overcoming traditional IL preprocessing challenges [1, 3, 8]. HIL enables robots to learn from expert demonstrations through a two-tier policy structure: acquiring sub-policies for specific task segments at the lower level and overarching strategies for skill transition at the higher level. However, HIL's effectiveness depends on the robustness of its hierarchical structure, with weaknesses leading to subpar imitation. Recognizing the limitations of relying solely on deep learning for hierarchy development in HIL, we draw inspiration from human cognitive processes in skill acquisition. This process emphasizes the dynamic interaction of information processing, task decomposition, decision-making, and refinement, with a significant emphasis on the integration of prior knowledge and observed behaviors through working memory [4, 6, 7]. Building on these principles, we propose the Cognition-Action-based Robotic Skill Imitation Learning (CasIL) framework. CasIL introduces a novel dual cognition-action structure for effective skill imitation in complex tasks, incorporating operators' cognitive priors for enhanced learning efficiency.

2 COGNITION-ACTION-BASED SKILL IMITATION LEARNING

In our problem formulation, we model long-horizon task environments as Semi-Markov Decision Processes (SMDP), represented by the tuple $(\mathcal{S}, \mathcal{A}, \{\mathcal{I}_o, \pi_o, \beta_o\}_{o\in\mathcal{O}}, \pi_{\mathcal{O}}(o|s), \mathcal{P}, \mathcal{R})$. Here, $\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}$ are standard MDP components, with the addition of $\{\mathcal{I}_o, \pi_o, \beta_o\}$ for each option o in the option set $\mathcal{O}$. An option, comprising a policy $\pi_o : \mathcal{S}\times\mathcal{A} \to [0,1]$, a termination condition $\beta_o : \mathcal{S}^+ \to [0,1]$, and an initiation set $\mathcal{I}_o \subseteq \mathcal{S}$, is valid in state s_t iff $s_t \in \mathcal{I}_o$. The system transitions between options based on the termination condition β_o and the inter-option policy $\pi_{\mathcal{O}}(o|s)$, progressing until the task is completed. The CasIL framework, illustrated in Fig. 1, features three main components. Initially, pre-trained image and text encoders process the visual and textual inputs. Following this, a cognition generator $\mathcal{F} : G\times\mathcal{S} \to \mathcal{O}$ and a policy module $\pi_{\mathcal{O}} : \mathcal{S}\times\mathcal{O} \to \mathcal{A}$ operate in tandem. Here, $\mathcal{O} = \{o^{\mathbf{1}}, \dots, o^{\mathbf{K}}\}$ denotes a set of K options, with each option representing a sub-task equipped with a specific skill, together forming a skill chain. CasIL works through two phases: 1) Leveraging manually inputted cognitive priors for task decomposition and expert visual demonstrations, the robot constructs its

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

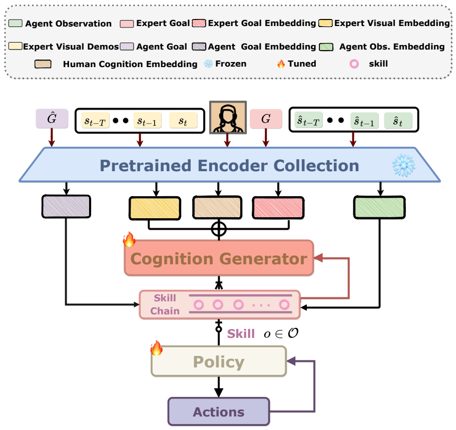

Figure 1: The workflow of CasIL.

cognition-action framework and skill chain $\mathcal{O}$, guided by the task objectives. 2) The robot then chooses the most appropriate sub-task skill from $\mathcal{O}$, based on its observation history, and learns and implements the relevant policies $\pi_{\mathcal{O}}$ to accomplish the sub-tasks. Using expert demonstration $\tau^E = (G, \{s_t, a_t\}_{t=1}^{T})$ and textual decompositions $\{l_{\mathbf{1}}, \ldots, l_{\mathbf{K}}\}$ based on human cognitive priors, the cognition generator $\mathcal{F}$ aligns states s_t with decompositions $l_{\mathbf{t}}$ to produce essential skills $o^{\mathbf{t}}$ for each division, extending the demonstration into an option-expanded trajectory $\tau^E = (G, \{s_t, o_{\mathbf{t}}, a_t\}_{1\leq t\leq T}^{\mathbf{1}\leq \mathbf{t}\leq \mathbf{K}})$, creating an SMDP structure. The robot selects relevant skills $o^{\mathbf{t}}$ based on the goal and observations, with each skill $o^{\mathbf{t}}$ active for $H(\mathbf{t})$ time steps. The policy π then guides actions at each step, depending on the state and the current skill. A CasIL-equipped robot utilizes human cognitive priors to learn and form its cognition of skills from expert demonstrations, focusing on critical decision-making steps. This learning approach enables the robot to adapt its actions based on observed inputs, following a dual learning framework. CasIL's training involves both a high-level cognition generator for skill chain encoding and a low-level action module for skill execution, as illustrated in Fig. 1. The cognition generator aligns task goals with human cognitive priors and expert demonstrations, while the low-level module employs behavior cloning with options. The training objective for a trajectory of length T aims to minimize the loss function:

$$\mathcal{L}_{\text{CasIL}} = \min_{\theta_g, \theta_p} \sum_{t=1}^{T} \Big(- \varepsilon \log \mathcal{F}_{\xi}(o^{\mathbf{t}} | G, \{s_\kappa\}_{\kappa=1}^{\kappa=\sum H(\mathbf{t})}, \{o^\kappa\}_{\kappa=\mathbf{1}}^{\mathbf{t}-\mathbf{1}}) \quad (1)$$
$$\log \pi_\theta(\hat{a}_t | \hat{G}, \{\hat{s}_\kappa\}_{\kappa=1}^{\kappa=\sum H(\mathbf{t})}, o^{\mathbf{t}}) \Big).$$

where ξ and θ represent the weights for the cognition generator and policy module, respectively, and ε adjusts the cognitive generation loss. The symbols o, s, a, and G denote skills, observations, actions, and task goals, consistent with the initial definitions.

3 EXPERIMENTS

Our experiments on RLBench [2] assess the methods in robotic arm manipulation tasks across four increasingly complex settings, each with 100 demonstration trajectories for training. Each setup involves a 6-DOF robotic arm with a gripper: **ToiletSeatDown:** The task is to lower the toilet lid onto the seat within 200 time steps. **PutRubbishInBin:** The robot must pick up and dispose of rubbish into a bin within 250 time steps. **PlayJenga:** The robot aims to remove a protruding block from a Jenga tower without toppling it, within 300 time steps. **InsertUsbInComputer:** The robot needs to pick up a USB stick and insert it into a USB port within 400 time steps. We assess models using success rates' mean and standard deviation in 80 randomized scenarios. Comparative methods include: **1) Supervised Behavioral Cloning (BC)** [5]: Lacks hierarchical structure and cognitive inputs. **2) Hierarchical Behavioral Cloning (H-BC)** [9]: Uses an option-based architecture without human cognitive priors. **3) Option-GAIL** [3]: Hierarchical, includes self-exploration but omits human cognitive guidance. **4) CasIL w/o Cognition**: CasIL variant without the cognition generator to highlight the importance of cognitive modeling. Test results in Table 1 reveal that all methods, including our CasIL, perform well in the simple *ToiletSeatDown* task, with CasIL achieving a 100% success rate across all test tasks. However, as task complexity increases (with more objects, longer periods and reduced stability), BC's success rate in skill imitation plummets, dropping to 0% in all *InsertUsbInComputer* test tasks. Baselines like H-BC and Option-GAIL, which lack the guidance of human cognitive priors, significantly lag behind CasIL in skill imitation. Similarly, CasIL w/o Cognition struggles with stable manipulation due to the absence of ongoing text-image alignment training. The performance of Option-GAIL, in particular, indicates that a one-step option architecture based solely on agent self-exploration fails to ensure stable skill imitation in long-horizon tasks.

	Robotic Arm Manipulation			
	ToiletSeatDown	*PutRubbishInBin*	*PlayJenga*	*InsertUsbInComputer*
BC	93.7 ± 4.3	74.4 ± 3.7	21.5 ± 8.8	00.0 ± 0.0
H-BC	98.5 ± 1.5	85.2 ± 5.9	33.6 ± 7.9	10.6 ± 1.8
Option-GAIL	99.0 ± 1.0	81.4 ± 9.6	48.2 ± 9.2	23.3 ± 5.5
CasIL w/o Cognition	99.4 ± 0.6	89.6 ± 9.4	53.1 ± 8.3	26.4 ± 4.1
CasIL (ours)	**100.0** ± **0.0**	**98.4** ± **1.6**	**82.4** ± **3.5**	**57.6** ± **2.4**

Table 1: Comparison of test results under four RLBench tasks.

4 CONCLUSION

We present CasIL, a framework for robot skill imitation using a dual cognition-action architecture. The framework utilizes a text-image-aligned skill chain that is derived from visual expert demonstrations and references human cognitive priors with manual input. This design facilitates robots in cognizing and imitating critical skills for long-horizon tasks. Experimental results show that CasIL improves robot skill imitation performance in long-horizon tasks. Future directions include further enriching cognitive priors and extending the task applicability of CasIL.

ACKNOWLEDGMENTS

This work was supported in part by the Science and Technology Innovation 2030 New Generation Artificial Intelligence Major Project under Grant 2021ZD0113303; in part by the National Natural Science Foundation of China under Grant 62192783, Grant 62276128; in part by the Collaborative Innovation Center of Novel Software Technology and Industrialization.

REFERENCES

[1] Jiayu Chen, Tian Lan, and Vaneet Aggarwal. 2023. Option-Aware Adversarial Inverse Reinforcement Learning for Robotic Control. In *IEEE International Conference on Robotics and Automation, ICRA 2023, London, UK, May 29 - June 2, 2023*. IEEE, 5902–5908.

[2] Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J Davison. 2020. Rlbench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters* 5, 2 (2020), 3019–3026.

[3] Mingxuan Jing, Wenbing Huang, Fuchun Sun, Xiaojian Ma, Tao Kong, Chuang Gan, and Lei Li. 2021. Adversarial option-aware hierarchical imitation learning. In *International Conference on Machine Learning*. PMLR, 5097–5106.

[4] Andrew N Meltzoff and Rebecca A Williamson. 2013. Imitation: Social, cognitive, and theoretical perspectives. (2013).

[5] Dean A Pomerleau. 1988. Alvinn: An autonomous land vehicle in a neural network. *Advances in neural information processing systems* 1 (1988).

[6] Yingxu Wang. 2007. On the Cognitive Processes of Human Perception with Emotions, Motivations, and Attitudes. *Int. J. Cogn. Informatics Nat. Intell.* 1, 4 (2007), 1–13.

[7] Yingxu Wang and Vincent Chiew. 2010. On the cognitive process of human problem solving. *Cogn. Syst. Res.* 11, 1 (2010), 81–92.

[8] Dandan Zhang, Qiang Li, Yu Zheng, Lei Wei, Dongsheng Zhang, and Zhengyou Zhang. 2021. Explainable hierarchical imitation learning for robotic drink pouring. *IEEE Transactions on Automation Science and Engineering* 19, 4 (2021), 3871–3887.

[9] Zhiyu Zhang and Ioannis Paschalidis. 2021. Provable hierarchical imitation learning via em. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 883–891.