A Robot Learning System for Viewpoint-aware Legible Motion Planning

Khai Nguyen*, Yi-Hung Chiu*, Prakrit Tyagi*, Saurav Kambil* and Inseung Kang
Department of Mechanical Engineering, Carnegie Mellon University
Email: {xuankhan, yihungc, prakritt, skambil, inseungk}@andrew.cmu.edu

Abstract—**Legibility is crucial for efficiency and trust in assistive robotics and shared autonomy, where the robot clearly conveys its goal through its motion. Previous planning techniques encounter challenges such as high computational latency, ambiguous objectives, and intensive tuning efforts. To overcome these obstacles, we propose a novel robot learning system that integrates the observer's viewpoint to generate adaptive, legible motions. We then demonstrate the effectiveness of two main learning paradigms: reinforcement learning and imitation learning. Our framework is validated through a goal-reaching manipulation task with an xArm6 robot in both simulated and real-world settings. Human evaluations indicate that our RL agent achieves a legibility score 15% higher than recorded demonstrations. Moreover, ablation studies show significant ambiguity when the robot fails to adapt to the observer's viewpoint.**

I. Introduction

Legible motion planning is crucial for enhancing assistive robotics and shared autonomy, particularly in environments where robots collaborate closely with humans, such as collaborative manufacturing and healthcare [1]. It involves designing robotic motions that are legible or *intent-expressive*, allowing humans to intuitively predict the robot's goal.

Learning-based approaches are increasingly favored for their scalability, robustness, and runtime efficiency compared to traditional methods like trajectory optimization [1]. Two main robot learning paradigms proven effective in achieving legibility are reinforcement learning (RL) [2] and imitation learning (IL) [3]. However, there has been limited research on leveraging the (human) observer model to improve legibility within learning-based techniques [4]. Inspired by the findings of Nikolaidis et al. [5], where *the viewpoint, as part of the observer model, plays a crucial role in legibility optimization*, we incorporate this insight into a learning-based planning system to better align robotic behavior with human perceptions. Our main contributions include:

- A novel robot learning system that effectively expresses intents by generating adaptive, legible motions considering the observer's viewpoint.
- Demonstrations of both RL and IL paradigms in accomplishing a goal-reaching manipulation task.
- Experiments conducted with the xArm6 robot in both simulated and real-world environments, followed by human evaluations to compare legibility scores.

The authors would like to thank Prof. Zackory Erickson for allowing us to use the xArm6 robot.

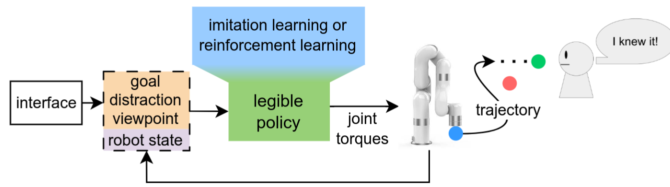

Fig. 1. Overview of our proposed robot learning system.

II. Proposed Robot Learning System

We propose a novel robot learning system that generates adaptive, legible motions by considering the observer's viewpoint (Fig. 1). In our goal-reaching manipulation task, the robot must move in a way that allows observers to easily recognize which of two objects (goal or distraction) it intends to reach. Moreover, the robot's actions need to adapt based on the observer's point-of-view, such as a top or side view. The policy's observation includes (extrinsic) task and observer model information (goal position, distraction position, and viewpoint input) along with the intrinsic robot state, enabling the agent to infer the appropriate action. In our experiments, we assume the robot knows its own state and receives extrinsic inputs from straightforward user interfaces, though these could also come from semantic perception [6] or vision-language models [7] for full autonomy. Our architecture can apply to both RL and IL setups, differing primarily in the specifics of the policy learning pipeline, detailed below.

A. Reinforcement Learning Pipeline

Reinforcement learning, characterized by a trial-and-error approach, allows robots to learn optimal actions through interactions with their environment, guided by a reward function that incentivizes legible behaviors.

Legibility Reward. We define a reward function:

$$r = -\alpha_1 d_{\text{goal}} + \alpha_2 g(d_{\text{distract}}, \beta_1) - \alpha_3 e_{\text{rot}} + \alpha_4 g(z, \beta_2). \quad (1)$$

$$g(x, y) = -1 \text{ if } x \leq y, \text{ otherwise } 0. \quad (2)$$

where the first term encourages the robot to reach the goal, the second term keeps it away from the distraction region (radius β_1), the third term maintains gripper orientation consistency, and the last term ensures it moves at least β_2 meters above the table. Importantly, this reward is computed in a 2D coordinate system corresponding to the side or top viewpoint. The task is considered successful when the robot reaches the goal.

Training Procedure. We represent the RL policy using a simple multi-layer perceptron (MLP) network. We randomly

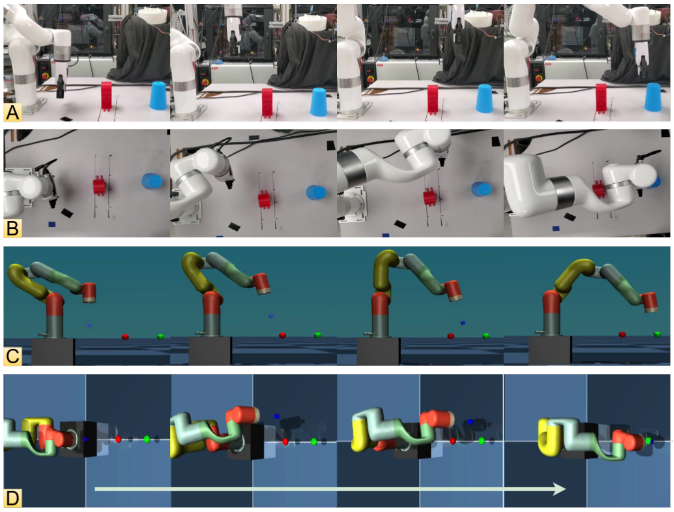

Fig. 2. Snapshots of the learned policies for imitation learning in real world and reinforcement learning in simulation. Our policy is able to generate legible plans for two views, side view (a) and (c), and top view (b) and (d).

sample goal and distraction positions within specified regions, along with a binary viewpoint (top view or side view). These inputs are concatenated with the robot state to create an observation, from which the RL agent infers its action in the form of joint torques. We use MuJoCo [8] as the physics simulator and the soft actor-critic (SAC) algorithm [9] to optimize the network's parameters. Policies with different seeds converge after roughly 200K training steps.

B. Imitation Learning Pipeline

In its simplest form, imitation learning, also known as behavior cloning, involves training a robot to mimic demonstrated behaviors by directly mapping observed states to corresponding actions, bypassing hard-to-specify task objectives.

Legible Human Demonstrations. A single human operator manually controls the robot to reach the goal in a way they consider legible. These observation-action pairs are recorded as human demonstrations or dataset. We found that a collection of 60 demonstrations, with various goal and distraction positions for two viewpoints, is sufficient to train the policy.

Training Procedure. We represent the IL policy using a simple MLP network and perform standard supervised learning on the collected legible observation-action data. The policy fits the training dataset within 1000 iterations.

III. Preliminary Results

A. Policy Learning

Our results for both RL and IL pipelines, in both simulation and real-world settings, highlight the critical role of the observer's viewpoint in effective, legible motion planning (Fig. 2). When observed from a side view, the robot must execute a vertical ascent to circumvent the distraction object, effectively expressing its intent. If viewed from the top, this behavior becomes intuitively ambiguous as the robot appears to head directly towards the distraction. To maintain legibility from this viewpoint, the learned policy navigates around the distraction object on the table plane.

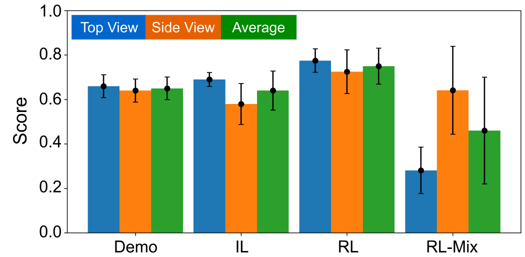

Fig. 3. Human evaluation results on the videos of different robot policies.

B. Human Evaluations

The RL reward function and IL demonstrations only approximate true legibility. To objectively assess our policies, we conducted a human study based on videos of human demonstrations, RL, and IL policies. We also tested an ablated scenario, RL-Mix (Fig. 3), to emphasize viewpoint adaptation. In RL-Mix, the RL policy was queried with one viewpoint but observed from another. Ten participants viewed all eight videos (two per policy, with four shown in Fig. 2). Participants paused the video when confident about the robot's intention and reported which object they thought the robot was targeting. We developed a legibility score system as follows:

$$\text{score} = \frac{t_{\text{video}} - t_{\text{pause}}}{t_{\text{video}}} \times \mathtt{check}(\text{guess}), \tag{3}$$

Comparing the scores in (Fig. 3), the IL-trained agent achieved nearly the same average score as its human demonstration counterpart, with better performance in the top view. Notably, the RL policy scored approximately 15% higher than human demonstrations. In the RL-Mix scenario, the top view video scored poorly at 0.281 because the robot appeared to head towards the distraction, resulting in longer pause times and incorrect answers. Moreover, individual perception variability is evident from the high standard error in the RL-Mix, particularly in the side view. Overall, the human evaluation shows that our robot learning system can generate adaptive, legible motion plans that align with human perceptions.

IV. Discussion and Future Work

Environment understanding boosts legibility. Our experiments show that the legibility of trajectories can vary significantly depending on the observer's viewpoint, highlighting its crucial role in decision-making processes. Our model inherently supports continuous parameterization of viewpoints.

Legibility is subjective and requires proper evaluation. On average, our RL policy outperforms human demonstrations as it optimizes behaviors based on carefully engineered rewards, whereas IL learns implicit objectives from data but is constrained by its quality. Our framework provides a foundation for evaluating learned legibility, with both RL and IL techniques objectively assessed in a unified context.

Future Directions. Our next steps involve evaluating the framework in more complex tasks involving multiple dynamic objects, comparing against established legible motion planning benchmarks [10]. We also aim to bridge the sim-to-real gap for RL and integrate richer sensory inputs such as perception to enhance adaptability and enable full autonomy.

References

[1] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. Legibility and predictability of robot motion. *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, pages 301–308, 2013.

[2] Baptiste Busch, Jonathan Grizou, Manuel Lopes, and Freek Stulp. Learning legible motion from human–robot interactions. *International Journal of Social Robotics*, 9 (5):765–779, 2017.

[3] Sebastian Wallkötter, Mohamed Chetouani, and Ginevra Castellano. Slot-v: supervised learning of observer models for legible robot motion planning in manipulation. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1421–1428. IEEE, 2022.

[4] Tathagata Chakraborti, Anagha Kulkarni, Sarath Sreedharan, David E Smith, and Subbarao Kambhampati. Explicability? legibility? predictability? transparency? privacy? security? the emerging landscape of interpretable agent behavior. In *Proceedings of the international conference on automated planning and scheduling*, volume 29, pages 86–96, 2019.

[5] Stefanos Nikolaidis, Anca Dragan, and Siddhartha Srinivasa. Viewpoint-based legibility optimization. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 271–278, 2016. doi: 10.1109/HRI.2016.7451762.

[6] Sourav Garg, Niko Sünderhauf, Feras Dayoub, Douglas Morrison, Akansel Cosgun, Gustavo Carneiro, Qi Wu, Tat-Jun Chin, Ian Reid, Stephen Gould, Peter Corke, and Michael Milford. Semantics for robotic mapping, perception and interaction: A survey. *Foundations and Trends® in Robotics*, 8(1–2):1–224, 2020.

[7] David Sobrín-Hidalgo, Miguel Ángel González-Santamarta, Ángel Manuel Guerrero-Higueras, Francisco Javier Rodríguez-Lera, and Vicente Matellán-Olivera. Enhancing robot explanation capabilities through vision-language models: a preliminary study by interpreting visual inputs for improved human-robot interaction, 2024.

[8] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033, 2012. doi: 10.1109/IROS.2012.6386109.

[9] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR, 2018.

[10] Mengxi Zhao, Rahul Shome, Ilana Yochelson, Kostas Bekris, and Eileen Kowler. An experimental study for identifying features of legible manipulator paths. In *International Symposium on Experimental Robotics*. Springer, 2014.