# Daesol Cho

*Research Statement*

(+82) 10-4748-0477
dscho1234@snu.ac.kr
dscho1234.github.io

## Research Topic

My research interest in robotics and deep reinforcement learning has been driven by the vision of creating autonomous systems capable of learning with minimal human intervention. Deep RL has enabled interactive agents to learn challenging skills in various domains and provided a versatile platform for robot learning, dealing with high-dimensional and multiple input modalities. However, there is still a gap between the promise of RL for autonomous robot learning and the reality of applying it in the real world. Humans still have to define some reward functions for tasks that require complex behavior or provide external interventions such as resets to the initial state due to the innate repetitive episodic nature of the RL framework. All of these are a major bottleneck for autonomy. This motivates a **RL method that minimizes human interventions which is the main theme of my doctoral research.**

Under the guidance of my advisor H. Jin Kim, I have focused on developing the RL framework, trying to eliminate or reduce human interventions by leveraging curriculum learning, data-driven reward, and a reset-free learning framework. Throughout my doctoral work, I have contributed to minimal intervention RL and demonstrated its potential on real robots through simulation and real-world hardware demonstration. My work can be characterized by the following key features: (i) learning without explicit rewards (**reward-free**), (ii) curriculum learning, and (iii) non-episodic training (**reset-free**).

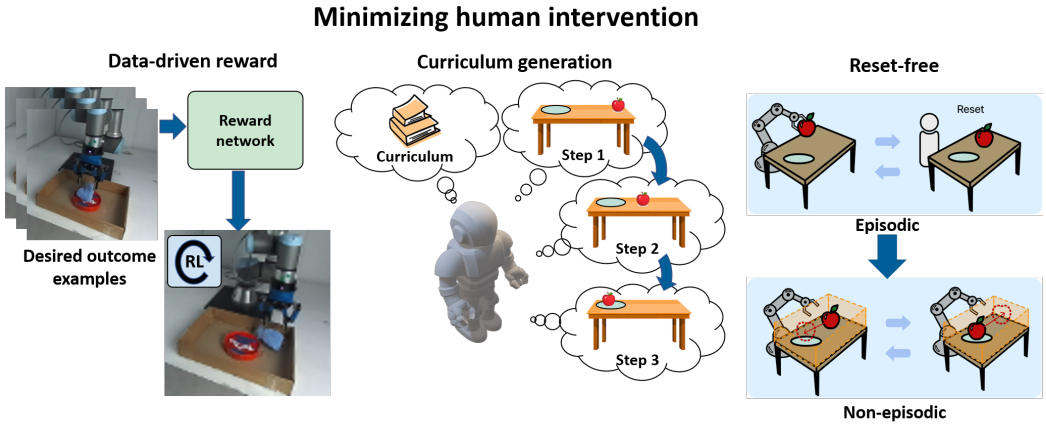



**Figure 1:** Minimizing external interventions for RL on embodied agents.

## Research Experience & Skills

Different intervention mitigation strategies are required depending on the purpose of the learning. To reduce efforts for reward specification and overall training setup, I have proposed ***OUTPACE*** [1] and ***D2C*** [2]. Both methods introduce a classifier-based learned reward and curriculum proposal mechanism, which only requires some desired outcome examples rather than full demonstration data, or privileged knowledge about the task. Also, to reduce the manual reset process in the real-world embodied agent, I have proposed ***IBC*** [3]. It introduces an additional RL agent to incorporate resets into the learning process which in effect divides a single continuous interaction into multiple forward and backward episodes. To further extend the ARL toward image inputs, I have also proposed ***A2TP*** (under review), which extends ARL with a more practical setup. It leverages readily available action-free video data, which extends the prior ARL work to the image inputs, and data-driven reward setting.

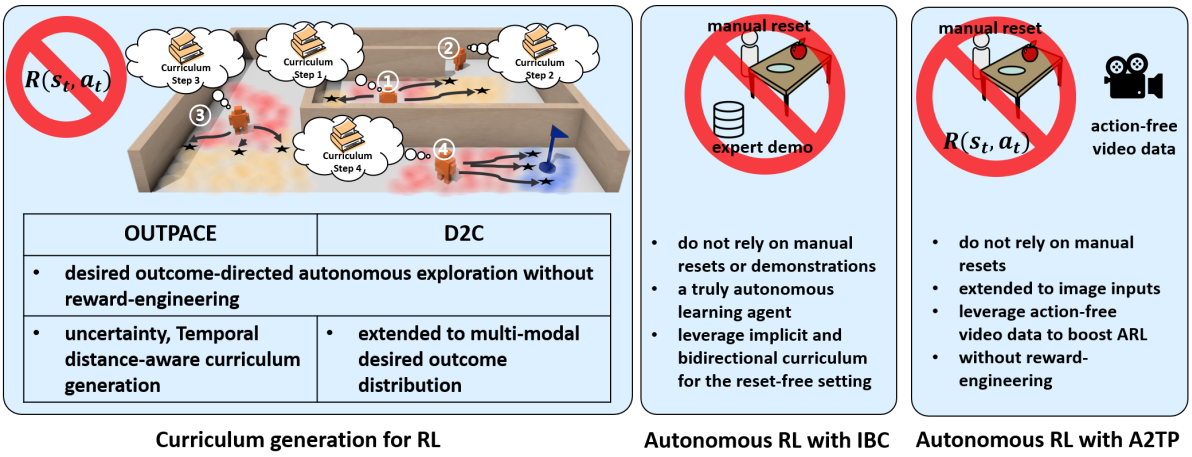



**Figure 2:** Summary of contributions.

*A. Outcome-directed RL via curriculum goal generation*

***OUTPACE*** trains a temporal-distance-aware function based on the Wasserstein distance with timestep metric. It is used to define the reward function, which is monotonically increasing along the optimal goal-reaching trajectory, providing a shaped reward for goal-reaching tasks. Along with a meta-learning-based classifier that discriminates already visited states and desired outcome states, it proposes curriculum goals gradually converging to the desired outcomes as learning progresses. It enables autonomous exploration of the environment by leveraging curriculum and learned reward function. Figure 3 visualizes the overview of the curriculum proposal process based on the proposed components.

As a follow-up work of ***OUTPACE***, ***D2C*** learns multiple classifiers trained to discriminate already visited states and desired outcome states while minimizing the mutual information between the classifiers with respect to the unexplored states. Unlike ***OUTPACE***, it does not depend on meta-learning-based classifiers, resulting in much faster inference and supporting multi-modal desired outcomes. These classifiers are used for both curriculum proposal and reward function, which enables autonomous learning of goal-reaching skills and exploration without human intervention. Figure 4 visualizes the overview of the curriculum proposal process based on the

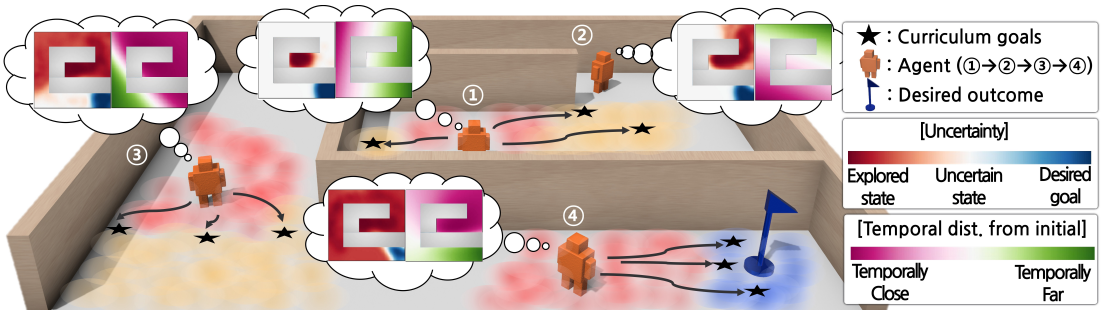



**Figure 3:** OUTPACE proposes uncertainty and temporal distance-aware curriculum goals to enable the agent to progress toward the desired outcome state automatically.

proposed components.

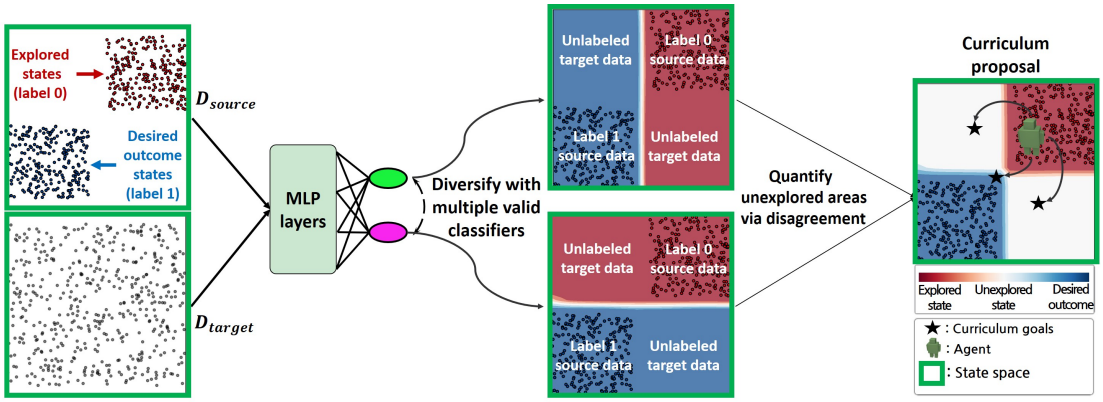


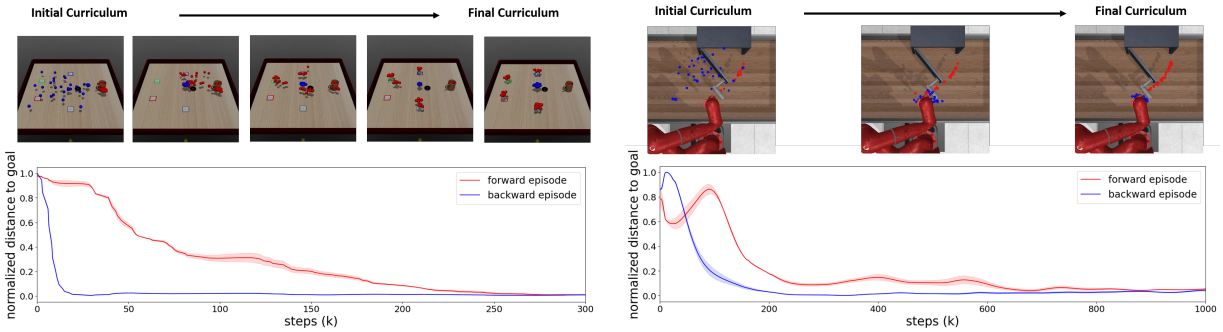

**Figure 4:** D2C trains a set of classifiers with labeled source data while diversifying their outputs on unlabeled target data (red: predicted label 0, blue: predicted label 1). Then, it proposes curriculum goals based on the diversified classifier's disagreement and similarity-to-desired outcome.

### *B. Autonomous (reset-free) RL*

Unlike previous autonomous RL methods, ***IBC*** based on implicit and bidirectional curriculum enables non-episodic acquisition of episodic skills without any manual resets and prior data. ***IBC*** consistently outperforms previous methods—even the ones that rely on prior data—both in terms of sample efficiency and final average success rate to achieve state-of-the-art performance. Conditional activation of the auxiliary agent implicitly shapes the effective initial state distribution of the forward agent and a bidirectional goal curriculum scheme based on the Wasserstein metric proposes goals of intermediate difficulty for both agents. Figure 5 visualizes curriculum goals and how they guide the learning process throughout non-episodic training.

### *C. Other RL Applications with Machine Learning Techniques*

My other work including research collaborations with colleagues focuses on applying machine learning to broad topics in robotics. I have proposed a mutual information-based unsupervised skill discovery method [4] for robot manipulation and have also proposed a data augmentation technique




**(a)** Tabletop Manipulation environment **(b)** Sawyer Door environment

**Figure 5:** ***IBC*** curriculum goals for the forward and auxiliary agents.

based on the generative model, GAN, for image-based offline RL [5]. In terms of curriculum learning and automating the reset process, I have also participated in Vector Quantization-based curriculum learning research [6] and self-supervised manual reset minimization research [7]. I co-developed the OpenAI Gym interface and MuJoCo models for physical systems (UR3 robot) in the process.

*D. Industry/Government Projects*

In addition to my research, I have participated in various industry-funded and government-funded projects implementing RL algorithms for robotic systems including 6DoF manipulators and automobiles as well as industrial control systems such as A/C. I had to integrate multiple models and frameworks, mostly between the simulated environment (MATLAB, Simulink) and the learning algorithm (PyTorch, TensorFlow), and developed technical expertise throughout the process. In addition to technical skills, I was able to learn interpersonal skills as I learned to effectively communicate with various researchers and engineers at different levels of expertise to set research objectives and reach a consensus.

## Future Research

While my research background is mostly in RL and robotics, I am also interested and excited to use a broad range of cutting-edge machine-learning tools, including diffusion models for complex multimodal robotic behavior synthesis and a Vision-Language-Model (VLM) for a user-friendly interface for controlling the robotic system. Building upon my doctoral research, I'd like to explore ways to apply generative/foundational models for autonomous agents. For example, diffusion models have the potential to introduce new types of policy or trajectory generators for complex multimodal behaviors in various applications. Also, VLM has the potential to provide new types of curriculum instructions or interventions which are much less costly to provide than domain-knowledge-intensive curriculum objectives or manual resets. I think these state-of-the-art machine learning tools can advance and expand the scope of autonomous learning even further.

## References

[1] D. Cho, S. Lee, and H. J. Kim, "Outcome-directed reinforcement learning by uncertainty & temporal distance-aware curriculum goal generation," *arXiv preprint arXiv:2301.11741*, 2023.

[2] ——, "Diversify & conquer: Outcome-directed curriculum rl via out-of-distribution disagreement," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[3] J. Kim, D. Cho, and H. J. Kim, "Demonstration-free autonomous reinforcement learning via implicit and bidirectional curriculum," in *Fortieth International Conference on Machine Learning*, 2023.

[4] D. Cho, J. Kim, and H. J. Kim, "Unsupervised reinforcement learning for transferable manipulation skill discovery," *IEEE Robotics and Automation Letters*, 2022.

[5] D. Cho, D. Shim, and H. J. Kim, "S2p: State-conditioned image synthesis for data augmentation in offline reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 11 534–11 546, 2022.

[6] S. Lee, D. Cho, J. Park, and H. J. Kim, "Cqm: Curriculum reinforcement learning with a quantized world model," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[7] J. Kim, J. H. Park, D. Cho, and H. J. Kim, "Automating reinforcement learning with example-based resets," *IEEE Robotics and Automation Letters*, 2022.