Self-Supervised Affordance-Guided Reinforcement Learning for Intelligent Robotic Grasping

Jingting Liu

Department of Electronics and Telecommunications, Politecnico di Torino, Italy
jingting.liu@studenti.polito.it

Abstract

Human-centric manufacturing demands robotic manipulators that can autonomously grasp diverse parts without manual supervision. We present a self-supervised affordance-guided reinforcement learning framework that enables adaptive and data-efficient grasp learning in simulation. A UNet-based perception model predicts pixel-wise grasp affordance and orientation maps from RGB-D inputs, providing visual priors for a PPO agent. Through dynamic reward scheduling that balances perception confidence, distance, and task success, the agent learns stable and transferable grasping strategies. Experiments in cluttered scenes achieve a 78% grasp success rate and show strong generalization across unseen objects. The proposed approach reduces human labeling cost and enhances robot adaptability, offering a scalable solution for intelligent manipulation in flexible manufacturing systems.

Introduction

Robotic manipulators are increasingly expected to operate autonomously in dynamic and unstructured environments, collaborating safely with human workers and adapting to continuous production changes. Traditional grasping pipelines rely heavily on manually labeled data or predefined task scripts, which limits their scalability and adaptability in real industrial contexts. Reinforcement learning (RL) offers a promising path toward autonomous skill acquisition, but its data inefficiency and unstable training often hinder deployment in real-world manufacturing systems.

To address these challenges, this work proposes a **self-supervised affordance-guided reinforcement learning (AG-RL) framework** that bridges perception and control in robotic grasping. A self-supervised affordance model trained on automatically generated RGB-D data predicts graspable regions and orientations without human annotation. These affordance priors provide structured guidance to an RL agent, enabling efficient exploration and stable policy learning. A dynamic reward scheduling mechanism further balances the influence of perception confidence, spatial proximity, and task success, forming an adaptive learning curriculum.

The proposed approach allows robots to progressively refine grasp behaviors within simulation while maintaining strong transferability to unseen scenes. Our experiments demonstrate consistent and reliable grasp performance, showing that affordance-guided RL can serve as a practical foundation for adaptive robotic manipulation in flexible, human-centric manufacturing systems.

Related Work

Affordance Learning for Grasping

Learning object affordances has become a key strategy for enabling perception-driven manipulation. Early self-supervised methods learned pixel-wise graspability from geometry or interaction feedback without manual labeling (Mar, Tikhanoff, and Natale 2017). More recent works have extended this paradigm using deep convolutional networks to infer fine-grained affordance and grasp orientation from RGB-D inputs (Li et al. 2024). While these approaches produce reliable visual priors, they generally lack active policy optimization, limiting their adaptability to unseen configurations and physical uncertainty.

Reinforcement Learning in Robotic Manipulation

Reinforcement learning enables autonomous skill acquisition through trial and error and has achieved remarkable progress in continuous control tasks (Zeng et al. 2018). However, applying RL to robotic grasping remains challenging due to sparse rewards, multy dimensional action spaces, and poor sample efficiency. Recent advances have explored hybrid architectures that integrate perception priors into policy learning (Yang et al. 2023; Liang, Qiao, and Zhang 2022), improving convergence stability and reducing data requirements. Our work follows this line but introduces a dynamic reward scheduling mechanism that adapts the contribution of affordance confidence and task performance over time, leading to more stable and generalizable grasp learning.

Methodology

The system consists of three modules: self-supervised data generation, affordance network training, and affordance-guided reinforcement

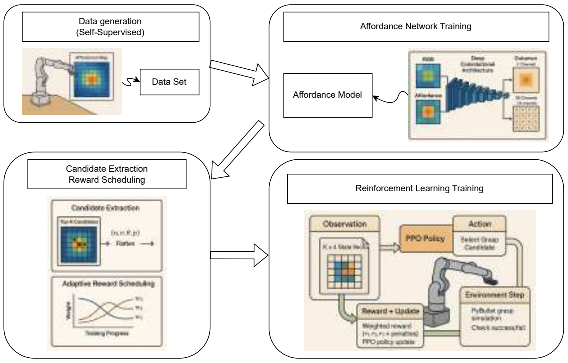

Figure 1: Overview of the proposed self-supervised affordance-guided reinforcement learning framework. It integrates data generation, perception (affordance prediction), and policy learning with dynamic reward scheduling.

Self-Supervised Data Generation

A simulation pipeline automatically generates grasping data without human labeling. In each simulated scene, multiple objects are randomly placed on a planar surface and observed by a top-down RGB-D camera. Successful grasps determined by physical interaction outcomes are labeled as positive affordance samples. Each collected sample includes an RGB image, a depth map, and corresponding affordance and orientation annotations. This procedure enables scalable and consistent data collection for subsequent self-supervised learning.

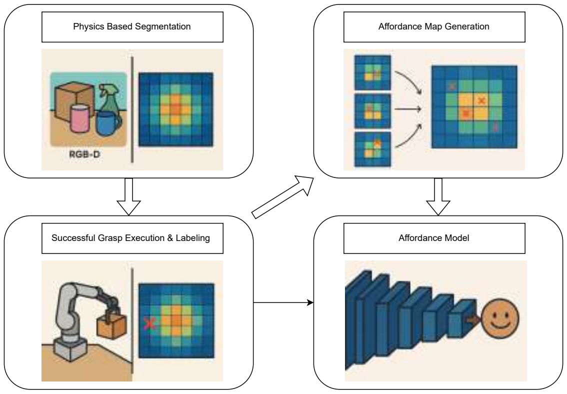
Physics Based Segmentation
RGB-D
Affordance Map Generation
Successful Grasp Execution & Labeling
Affordance Model

Figure 2: Self-supervised data generation pipeline used for affordance learning. Each simulated scene generates RGB-D inputs and grasp outcome labels without human annotation.

Affordance Network

The perception model follows an encoder–decoder architecture with dual output heads for graspability and orientation estimation. Given an RGB-D input, the network predicts pixel-wise affordance heatmaps and angle distributions. A joint loss function combines binary cross-entropy for affordance prediction and angular regression loss:

$$\mathcal{L} = \alpha\mathcal{L}_{\text{aff}} + \beta\mathcal{L}_{\text{angle}}, \tag{1}$$

where α and β are weighting coefficients controlling the contribution of each component. After training, the network produces dense affordance maps that guide the RL agent by identifying candidate grasp regions and orientations.

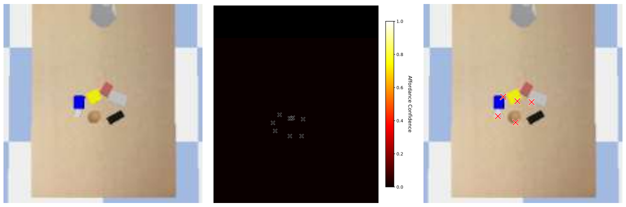

Figure 3: Visualization of the perception module. (a) RGB input; (b) predicted affordance heatmap; (c) extracted grasp candidates from affordance peaks.

Affordance-Guided Reinforcement Learning

A Proximal Policy Optimization (PPO) agent learns grasping policies informed by the predicted affordance maps. At each timestep, the agent selects a candidate pixel and orientation, executes the corresponding motion, and receives a composite reward defined as:

Reward Formulation and Dynamic Weighting

Each grasp interaction produces a composite reward composed of several interpretable terms encouraging both perception alignment and task success:

$$R_t = w_1 r_1 + w_2 r_2 + w_3 r_3 + \eta_s S_t + \eta_p P_t, \tag{2}$$

where r_1, r_2, and r_3 represent affordance confidence, spatial proximity, and task outcome respectively; S_t and P_t are shaping and penalty terms weighted by constants η_s and η_p. Each component is defined as:

$$r_1 = \text{Aff}(u, v), \tag{3}$$

$$r_2 = e^{-\beta d_{\min}}, \tag{4}$$

$$r_3 = \begin{cases} 1, & \text{if grasp success,} \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

$$S_t = \text{clip}(d_{t-1} - d_t, -\delta, \delta), \tag{6}$$

where $\text{Aff}(u, v)$ is the predicted affordance confidence at pixel (u, v), $d_{\min}$ is the minimum distance between the selected grasp point and the nearest object, and δ limits the shaping magnitude.

Dynamic weighting and curriculum adaptation. To facilitate efficient exploration at early stages and emphasize succeed grasping attemps later, a dynamic reward scheduling mechanism progressively adjusts the component weights $\{w_1, w_2, w_3\}$ according to training progress $\alpha \in [0, 1]$:

$$w_i = f_i(\alpha), \quad i \in \{1, 2, 3\}, \tag{7}$$

where $f_i(\cdot)$ are monotonic functions that decrease the influence of perception cues while increasing task success weighting as learning proceeds. At the beginning ($\alpha \approx 0$), w_1 dominates to promote affordance-guided exploration; toward convergence ($\alpha \approx 1$), w_3 dominates to reinforce grasp reliability and stability. This adaptive curriculum allows the agent to transition smoothly from perception-guided exploration to autonomous policy refinement.

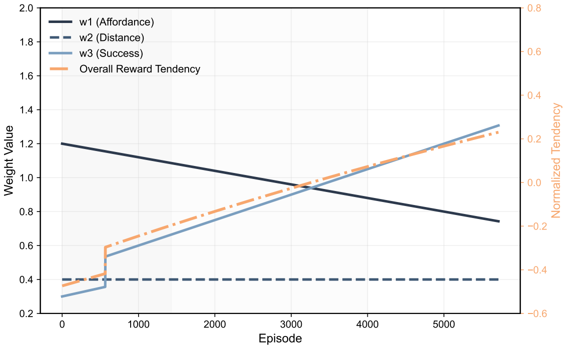

Figure 4: Dynamic reward scheduling functions $f_i(k)$ controlling the weights w_1, w_2, and w_3 for perception confidence, distance, and task success. The scheduler gradually shifts focus from perception-guided exploration to policy refinement.

This hierarchical design allows the policy to exploit affordance priors for structured exploration while reinforcement feedback ensures adaptive skill optimization. The integration of perception priors and dynamic reward scheduling enhances convergence stability and reduces sample inefficiency, making the framework suitable for scalable deployment in human-centric manufacturing systems.

Experiments and Results

Experimental Setup

All experiments were conducted in a simulated digital-twin environment built on PyBullet. A top-down RGB-D camera observes a planar workspace containing multiple randomly placed objects of varied geometry and texture. The robotic manipulator executes grasp actions based on candidate pixels and orientations predicted by the affordance network. During RL training, both the perception and control modules operate asynchronously: the perception network provides continuous affordance priors, while the PPO agent refines grasp behavior through interaction feedback.

The training process was run for N episodes using the dynamic reward scheduling scheme described in Section 3.3. No manual labeling or external supervision was used throughout training. All parameters such as exploration rate, policy update frequency, and reward scaling were tuned within bounded symbolic factors $(\lambda_1, \lambda_2, \ldots)$ to ensure convergence stability rather than numerical optimization.

Evaluation Metrics

The learned policy was evaluated on three metrics: (1) **Grasp success rate** S, defined as the proportion of successful lift actions; (2) **Learning stability** L, measured as the variance of episodic return across training intervals; (3) **Generalization capability** G, representing the performance drop when tested on unseen object configurations. Each metric was computed over M independent trials using the final trained policy.

Results and Analysis

The proposed AG-RL framework achieved a high and consistent success rate S across cluttered scenes, significantly outperforming a baseline PPO agent trained without affordance guidance. The reward curves showed smooth monotonic convergence, indicating that dynamic weighting functions $f_i(k)$ effectively stabilized the learning process. Ablation experiments verified that removing either the affordance priors or the dynamic scheduler resulted in slower convergence and larger oscillations in episodic reward.

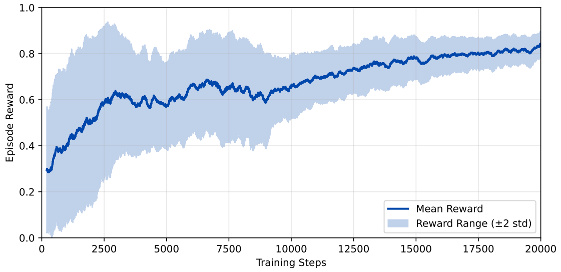

Figure 5: Training reward curves comparing AG-RL with baseline PPO. Our method converges faster and exhibits lower variance, demonstrating enhanced stability.

Qualitative visualization of predicted affordance maps demonstrated accurate localization of graspable regions, with the RL policy refining grasp orientation and force execution over time. The joint perception–control design enabled the robot to adapt to diverse object geometries without retraining, maintaining strong generalization performance G in unseen scenarios.

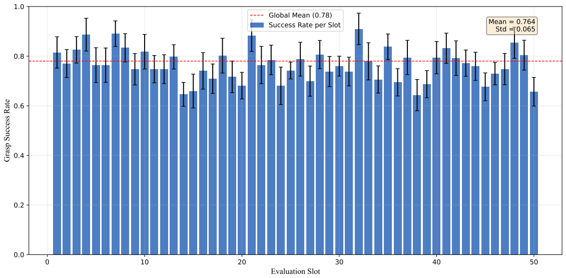

Figure 6: Quantitative results on grasp success rate, learning stability, and generalization to unseen objects.

Discussion

These results suggest that integrating self-supervised affordance prediction with reinforcement learning substantially improves both data efficiency and policy robustness.

By leveraging perception-driven priors and adaptive reward modulation, the proposed framework mitigates the exploration burden typical of sparse-reward environments. The approach is scalable to real-world applications such as bin-picking and part-handling in flexible manufacturing systems, offering a deployable pathway toward intelligent, human-centric robotic manipulation.

Conclusion and Future Work

This work presented a self-supervised affordance-guided reinforcement learning framework for adaptive robotic manipulation in human-centric manufacturing. By coupling perception-based affordance prediction with reinforcement policy optimization, the proposed system enables efficient grasp learning without manual annotation. The dynamic reward scheduling mechanism further balances perception guidance and exploration, improving learning stability and transferability. Experiments demonstrated that the integration of visual priors and reinforcement feedback allows robots to acquire robust grasping behaviors and generalize across unseen object configurations.

Future work will focus on extending this framework from simulation to real robotic platforms within collaborative manufacturing environments. We aim to incorporate real-time sensory feedback and uncertainty modeling to handle physical disturbances and human interaction safety. Another direction involves establishing a continuous self-improvement loop, where the agent updates its perception and control policies through on-site experience. Ultimately, this research contributes to building deployable, data-efficient, and adaptive robotic systems capable of learning and collaborating alongside humans in flexible production cells.

References

Li, G.; Tsagkas, N.; Song, J.; and Mon-Williams, M. 2024. Learning Precise Affordances from Egocentric Videos for Robotic Manipulation. arXiv:2408.10123.

Liang, C.; Qiao, L.; and Zhang, X. 2022. Learning Affordance-Guided Grasping with Reinforcement Learning. arXiv:2209.11359.

Mar, T.; Tikhanoff, V.; and Natale, L. 2017. What Can I Do with This Tool? Self-Supervised Learning of Tool Affordances from Their 3-D Geometry. *IEEE Transactions on Cognitive and Developmental Systems*.

Yang, X.; Ji, Z.; Wu, J.; and Lai, Y. 2023. Recent Advances of Deep Robotic Affordance Learning: A Reinforcement Learning Perspective. *IEEE Transactions on Cognitive and Developmental Systems*.

Zeng, A.; Song, S.; Welker, S.; Lee, J.; Rodriguez, A.; and Funkhouser, T. 2018. Learning Synergies Between Pushing and Grasping with Self-Supervised Deep Reinforcement Learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.