# Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes

**Chen Tang**[*1], **Ben Abbatematteo**[*1] , **Jiaheng Hu**[*1], **Rohan Chandra**[2],
**Roberto Martín-Martín**[1], **Peter Stone**[1,3]
[1] The University of Texas at Austin,
[2] The University of Virginia,
[3] Sony AI

## Abstract

Reinforcement learning (RL), particularly its combination with deep neural networks referred to as deep RL (DRL), has shown tremendous promise across a wide range of applications, suggesting its potential for enabling the development of sophisticated robotic behaviors. Robotics problems, however, pose fundamental difficulties for the application of RL, stemming from the complexity and cost of interacting with the physical world. This overview article provides a concise summary of our recently released survey paper "*Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes*", which has been accepted for publication in *Annual Review of Control, Robotics, and Autonomous Systems*. The survey paper provides a modern survey of DRL for robotics, with a particular focus on evaluating the real-world successes achieved with DRL in realizing several key robotic competencies. Our analysis aims to identify the key factors underlying those exciting successes, reveal underexplored areas, and provide an overall characterization of the status of DRL in robotics. We highlight several important avenues for future work, emphasizing the need for stable and sample-efficient real-world RL paradigms, holistic approaches for discovering and integrating various competencies to tackle complex long-horizon, open-world tasks, and principled development and evaluation procedures. The survey is designed to offer insights for both RL practitioners and roboticists toward harnessing RL's power to create generally capable real-world robotic systems.

**Full Version** — https://www.arxiv.org/abs/2408.03539

## Introduction

Reinforcement learning (RL) (Sutton and Barto 2018) refers to a class of decision-making problems in which an agent must learn through trial-and-error to act in such a way that maximizes its accumulated *return*, as encoded by a scalar reward function that maps the agent's states and actions to immediate rewards. RL algorithms, particularly their combination with deep neural networks referred to as deep RL (DRL) (François-Lavet et al. 2018), have shown remarkable capabilities in solving complex decision-making problems even with high-dimensional observations in domains such as board games (Schrittwieser et al. 2020), video games (Wurman et al. 2022), healthcare (Yu et al. 2021), and recommendation systems (Afsar, Crump, and Far 2022).

These successes underscore the potential of DRL for controlling robotic systems with high-dimensional state or observation spaces and highly nonlinear dynamics to perform challenging tasks that conventional decision-making, planning, and control approaches (e.g., classical control, optimal control, sampling-based planning) cannot handle effectively. Yet, the most notable milestones of DRL so far have been achieved in simulation or game environments, where RL agents can learn from extensive experience. In contrast, robots must complete tasks in the *physical world*, which presents additional challenges. It is often inefficient and/or unsafe for the RL agents to collect trial-and-error samples directly in the physical world, and it is usually impossible to create an exact replica of the complex real world in simulation. These challenges notwithstanding, recent advances have enabled DRL to succeed at some real-world robotic tasks. For instance, DRL has enabled champion-level drone racing (Kaufmann et al. 2023) and versatile quadruped locomotion control integrated into production-level quadruped systems (e.g., ANYbotics, Swiss-Mile, and Boston Dynamics). However, *the maturity of state-of-the-art DRL solutions varies significantly across different robotic applications*. In some domains, such as urban autonomous driving, DRL-based solutions remain limited to simulation or strictly confined cases (Kiran et al. 2021).

Our survey aims to comprehensively evaluate the current progress of DRL in real-world robotic applications, identifying key factors behind the most exciting successes and open challenges that remain in less mature areas. Specifically, we assess the maturity of DRL for a variety of problem domains and contrast the DRL literature across domains to pinpoint broadly applicable techniques, under-explored areas, and common open challenges that need to be addressed to advance DRL's applications in robotics. We aim for the survey to provide researchers and practitioners with a thorough understanding of the status of DRL in robotics, offering valuable insights to guide future research and facilitate broadly deployable DRL solutions for real-world robotic tasks.

[*] These authors contributed equally.

## Why Another Survey on RL for Robotics?

Although some previous articles have surveyed RL for robotics, we make three contributions that provide unique perspectives on the literature and fill gaps in knowledge. First, we focus on work that has demonstrated at least *some degree of real-world success*, aiming to assess the current state and open challenges of DRL for real-world robotic applications. Most existing surveys on RL for robotics do not explicitly address this topic, e.g., Dulac-Arnold et al. (2021) discuss the general challenges of real-world RL not specific to robotics, and Ibarz et al. (2021) list open challenges of DRL unique to real-world robotics settings but based on case studies drawn only from their own research. In contrast, our discussion is grounded in a comprehensive assessment of the real-world successes achieved by DRL, with one aspect of our evaluation being the level of real-world deployment.

Second, we present a *novel* and *comprehensive* taxonomy that categorizes DRL solutions along multiple axes: robot competencies learned with DRL, problem formulation, solution approach, and level of real-world success. Prior surveys on RL for robotics and broader robot learning have often focused on specific tasks (Kroemer, Niekum, and Konidaris 2021; Xiao et al. 2022) or on particular techniques (Deisenroth 2011; Brunke et al. 2022). By contrast, our taxonomy allows us to survey the complete landscape of DRL solutions that are effective in robotics application domains, in addition to reviewing the literature of each application domain separately. Within this framework, we compare and contrast solutions and identify *common patterns, broadly applicable approaches, under-explored areas, and open challenges* for realizing successful robotic systems.

Third, while some past surveys have shared our motivation to provide a broad analysis of the field, the fast and impressive pace of DRL progress has created the need for a renewed analysis of the field, its successes, and limitations. The seminal survey by Kober, Bagnell, and Peters (2013) was written before the deep learning era, and the general deep learning for robotics survey by Sünderhauf et al. (2018) was written when DRL accomplishments were primarily in simulation. We provide a refreshed overview of the field by focusing on DRL, which is behind the most notable real-world successes of RL in robotics, paying particular attention to papers published in the last five years.

## Taxonomy

This section presents the novel taxonomy we introduce to categorize the literature on DRL. The unique focus of our survey on the real-world successes of DRL in robotics necessitates a new taxonomy to categorize and analyze the literature, which should enable us to assess the maturity of DRL solutions across various robotic applications and derive valuable lessons from both successes and failures. Specifically, we should identify the specific robotic problem addressed in each paper, understand how it has been abstracted as an RL problem, and summarize the DRL techniques applied to solve it. More importantly, we should evaluate the maturity of these DRL solutions, as demonstrated in their experiments. Consequently, we introduce a taxonomy spanning four axes: **robot competencies learned with DRL**, **problem formulation**, **solution approach**, and **the level of real-world success**.

### Robot Competencies Learned with DRL

Our primary axis focuses on the target robotic task studied in each paper. A robotic task, especially in open real-world scenarios, may require multiple competencies. DRL can be applied to synthesize an end-to-end system to realize all the competencies or learn sub-modules to enable a subset of them. Since our focus is DRL, we classify papers based on *the specific robot competencies learned and realized with DRL*. At the highest level, we classify the competencies into *single-robot*—competencies required for a robot to complete tasks on its own—and *multi-agent*—competencies required to interact with other agents sharing the workspace with the robot and affecting its task completion.

When a single robot completes a task in a workspace, any competencies it requires can be considered as enabling specific ways to *interact with and affect the physical world*, which are further divided into `mobility`—moving in the environment—and `manipulation`—moving or rearranging (e.g., grasping, rotating) objects in the environment (Mason 2001; Siciliano, Khatib, and Kröger 2008; Mason 2018). In the robotics literature, mobility[1] is typically split into two problems: `locomotion` and `navigation` (Siciliano, Khatib, and Kröger 2008; Rudin et al. 2022). Locomotion focuses on motor skills that enable robots of various morphologies (e.g., quadrupeds, humanoids, wheeled robots, drones) to traverse different environments, while navigation focuses on strategies that direct a robot to its destination efficiently without collision. Typical navigation policies generate *high-level* motion commands, such as desired states at the center of mass (CoM), while assuming effective locomotion control to execute them (Siciliano, Khatib, and Kröger 2008). Some works jointly address the locomotion and navigation problems, which is particularly useful for tasks in which the navigation strategies are heavily affected by the robot's capability to traverse the environment, as determined by the robot dynamics and locomotion control. We review these papers alongside other navigation papers since their ultimate goal is navigation.

In the robotics literature, manipulation is often studied in table-top settings, e.g., robotic arms or hands mounted on a stationary base with fixed sensors observing the scene. Some other real-world tasks further require robots to interact with the environment while moving their base, which necessitates a synergistic integration of manipulation and mobility capabilities. We review the former case under the `stationary manipulation` category and the latter under `mobile manipulation`.

When the task completion is affected by the other agents in the workspace, the robot needs to be further equipped with

[1] In the robotics literature, both `locomotion` and `navigation` have been used to refer to the ability to move in an environment. To avoid confusion, `mobility` is used in this survey to refer to the overarching category where DRL enables robot movement.

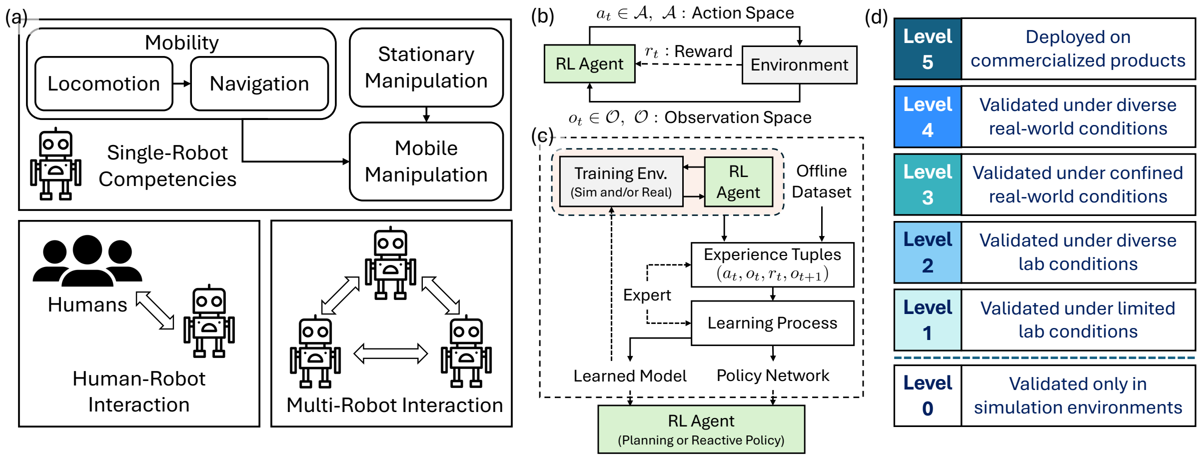


Figure 1: The four aspects of our taxonomy: (a) Robot competencies learned with DRL; (b) Problem formulation; (c) Solution approach; and (d) Levels of real-world success.

*abilities to interact with other agents*, which we place under the heading of *multi-agent* competencies. Note that some single-robot competencies may still be required while the robot interacts with others, such as crowd navigation or collaborative manipulation. In this category, we focus on papers where DRL occurs at the agent-interaction level, i.e., learning interaction strategies given certain single-robot competencies or learning policies that jointly optimize interaction and single-robot competencies. We further split these works into two subcategories based on the types of agents the robot interacts with: 1) *Human-robot interaction* concerns a robot's ability to operate alongside humans. The presence of humans introduces additional challenges due to their sophisticated behavior and the stringent safety requirements for robots operating around humans. 2) *Multi-robot interaction* refers to a robot's ability to interact with a group of robots. A class of RL algorithms, multi-agent RL (MARL), is typically applied to solve this problem. In MARL, each robot is a learning agent evolving its policy based on its interactions with the environment and other robots, which complicates the learning mechanism. Depending on whether the robots' objectives align, their interactions could be cooperative, adversarial, or general-sum. In addition, practical scenarios often require decentralized decision-making under partial observability and limited communication bandwidth.

## Problem Formulation

The second axis of our taxonomy is the formulation of the RL problem, which specifies the optimal control policy for the targeted robot competency. RL problems are typically modeled as Partially Observable Markov Decision Processes (POMDPs) for single-agent RL and Decentralized POMDPs for multi-agent RL. Specifically, we categorize the papers based on the following elements of the problem formulation: 1) *Action space*: whether the actions are *low-level* (i.e., joint or motor commands), *mid-level* (i.e., task-space commands), or *high-level* (i.e., temporally extended task-space commands or subroutines); 2) *Observation space*: whether the observations are *high-dimensional* sensor inputs (e.g., images and/or LiDAR scans) or estimated *low-dimensional* state vectors; 3) *Reward function*: whether the reward signals are *sparse* or *dense*.

## Solution Approach

Another axis closely related to the previous one is the solution approach used to solve the RL problem, which is composed of the RL algorithm and associated techniques that enable a practical solution for the target robotic problem. Specifically, we classify the solution approach from the following perspectives: 1) *Simulator usage*: whether and how simulators are used, categorized into *zero-shot*, *few-shot sim-to-real transfer*, or directly learning offline or in the real world *without simulators*; 2) *Model learning*: whether (a part of) the transition dynamics model is learned from robot data; 3) *Expert usage*: whether expert (e.g., human or oracle policy) data are used to facilitate learning; 4) *Policy optimization*: the policy optimization algorithm adopted, including *planning* or *offline*, *off-policy*, or *on-policy RL*; 5) *Policy/Model Representation*: Classes of neural network architectures used to represent the policy or dynamics model, including *MLP*, *CNN*, *RNN*, and *Transformer*.

## Level of Real-World Success

To evaluate the practicality of DRL in real-world robotic tasks, we categorize the papers based on the maturity of their DRL methods. By comparing the effectiveness of DRL across different robotic tasks, we aim to identify domains where the gaps between research prototypes and real-world deployment are more or less significant. This requires a metric to quantify real-world success across tasks, which, to our knowledge, has not been attempted in the DRL for robotics literature. Inspired by the levels of autonomous driving (On-Road Automated Driving (ORAD) committee 2018) and Technology readiness level (TRL) for machine learning (Lavin et al. 2022), we introduce the concept of *levels of real-world success*. We classify the papers into six levels based on the scenarios where the proposed methods have been validated: 1) *Level 0*: validated only in simulation; 2) *Level 1*: validated in limited lab conditions; 3) *Level 2*: validated in diverse lab conditions; 4) *Level 3*: validated under confined real-world operational conditions; 5) *Level 4*: validated under diverse, representative real-world operational conditions; and 6) *Level 5*: deployed on commercialized products. We consider Levels 1-5 as achieving at least some degree of real-world success. The only information we can use to assess the level of real-world success is the experiments reported by the authors. However, many papers only described a single real-world trial. While we strive to provide accurate estimates, this assessment can be subjective due to limited information. Additionally, we use the level of real-world success to quantify the maturity of a solution for its target problem, irrespective of its complexity.

## Review Overview

In our survey paper (Tang et al. 2024), we provide a detailed review of the literature, grouped by the robot competencies learned with DRL. For each robot competency, we further organize the review based on subcategories specific to each type of competency and discuss the papers based on our proposed taxonomy. We then summarize the trends and open challenges for learning each competency in question. Please refer to Section 4 of the survey paper (Tang et al. 2024) for the detailed review.

Here, we briefly summarize the general patterns we found behind current real-world successes in robotics achieved with DRL and the characteristics of those less successful cases. Overall, more mature solutions (i.e., L3-4) have often followed the zero-shot sim-to-real transfer scheme, which works particularly well for locomotion and navigation. The dynamics involved in these competencies, especially terrestrial locomotion and navigation, are relatively stable and easy to simulate. Dense and shaped rewards, which simplify exploration and improve sample efficiency, have also been effective, leading to the predominant use of stable and robust model-free, on-policy algorithms in these domains. The sim-to-real scheme has been successful for manipulation problems in which dense reward functions can be designed a priori (e.g., grasping, assembly, in-hand, non-prehensile manipulation), but less so in tasks with more diversity (e.g., pick-and-place). The community has been striving to explore alternative solutions that do not require simulation or reward shaping and adopt policy optimization algorithms with better sample efficiency. Human demonstrations are effective for enabling real-world learning, particularly in manipulation tasks that are not prohibitively complex to demonstrate. For competencies where both accurate simulation and real-world rollouts are prohibitive (e.g., HRI) or where stable, scalable RL algorithms are missing (e.g., multi-robot interaction), successful real-world examples are much sparser.

We also pinpoint several concrete open challenges that are opportunities for further extending DRL's applications, in particular for those currently less successful domains, including: 1) improving stability and sample-efficiency in RL algorithms; 2) real-world learning; 3) learning for long-horizon robotic tasks; 4) designing principled approaches for RL systems; 5) benchmarking real-world success; and 6) harnessing foundation models for DRL in robotics. Please refer to the survey paper (Tang et al. 2024) for a detailed discussion.

## Acknowledgements

We thank Pieter Abbeel, Jan Peters, Yuchen Cui, Shivin Dass, George Konidaris, Eric Rosen, Koushil Sreenath, Eugene Vinitsky, and Zhaoming Xie for their feedback. A portion of this work has taken place in the Learning Agents Research Group (LARG) at the Artificial Intelligence Laboratory at the University of Texas at Austin. LARG research is supported in part by the National Science Foundation (FAIN-2019844, NRT-2125858), the Office of Naval Research (N00014-18-2243), Army Research Office (E2061621), Bosch, Lockheed Martin, and Good Systems, a research grand challenge at the University of Texas at Austin. The views and conclusions contained in this document are those of the authors alone. Peter Stone serves as the Chief Scientist of Sony AI and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

## References

Afsar, M. M.; Crump, T.; and Far, B. 2022. Reinforcement learning based recommender systems: A survey. *ACM Comput. Surv.*, 55(7): 1–38.

Brunke, L.; Greeff, M.; Hall, A. W.; Yuan, Z.; Zhou, S.; Panerati, J.; and Schoellig, A. P. 2022. Safe Learning in Robotics: From Learning-Based Control to Safe Reinforcement Learning. *Annu. Rev. Control Robot. Auton. Syst.*, 5(1): 411–444. _eprint: https://doi.org/10.1146/annurev-control-042920-020211.

Deisenroth, M. P. 2011. A Survey on Policy Search for Robotics. *Found. Trends Robot.*, 2(1-2): 1–142.

Dulac-Arnold, G.; Levine, N.; Mankowitz, D. J.; Li, J.; Paduraru, C.; Gowal, S.; and Hester, T. 2021. Challenges of real-world reinforcement learning: definitions, benchmarks, and analysis. *Mach. Learn.*, 110(9): 2419—-2468.

François-Lavet, V.; Henderson, P.; Islam, R.; Bellemare, M. G.; Pineau, J.; et al. 2018. An introduction to deep reinforcement learning. *Found. Trends in Mach. Learn.*, 11(3-4): 219–354.

Ibarz, J.; Tan, J.; Finn, C.; Kalakrishnan, M.; Pastor, P.; and Levine, S. 2021. How to train your robot with deep reinforcement learning: lessons we have learned. *Int. J. Robot. Res.*, 40(4-5): 698–721.

Kaufmann, E.; Bauersfeld, L.; Loquercio, A.; Müller, M.; Koltun, V.; and Scaramuzza, D. 2023. Champion-level drone racing using deep reinforcement learning. *Nature*, 620(7976): 982–987.

Kiran, B. R.; Sobh, I.; Talpaert, V.; Mannion, P.; Al Sallab, A. A.; Yogamani, S.; and Pérez, P. 2021. Deep reinforcement learning for autonomous driving: A survey. *IEEE Trans. Intell. Transp. Syst.*, 23(6): 4909–4926.

Kober, J.; Bagnell, J. A.; and Peters, J. 2013. Reinforcement learning in robotics: A survey. *Int. J. Robot. Res.*, 32(11): 1238–1274.

Kroemer, O.; Niekum, S.; and Konidaris, G. 2021. A review of robot learning for manipulation: Challenges, representations, and algorithms. *J. Mach. Learn. Res.*, 22(30): 1–82.

Lavin, A.; Gilligan-Lee, C. M.; Visnjic, A.; Ganju, S.; Newman, D.; Ganguly, S.; Lange, D.; Baydin, A. G.; Sharma, A.; Gibson, A.; et al. 2022. Technology readiness levels for machine learning systems. *Nat. Commun.*, 13(1): 6039.

Mason, M. T. 2001. *Mechanics of robotic manipulation*. MIT press.

Mason, M. T. 2018. Toward robotic manipulation. *Annu. Rev. Control Robot. Auton. Syst.*, 1: 1–28.

On-Road Automated Driving (ORAD) committee. 2018. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Technical report, SAE International.

Rudin, N.; Hoeller, D.; Bjelonic, M.; and Hutter, M. 2022. Advanced skills by learning locomotion and local navigation end-to-end. In *IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2497–2503. IEEE.

Schrittwieser, J.; Antonoglou, I.; Hubert, T.; Simonyan, K.; Sifre, L.; Schmitt, S.; Guez, A.; Lockhart, E.; Hassabis, D.; Graepel, T.; et al. 2020. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839): 604–609.

Siciliano, B.; Khatib, O.; and Kröger, T. 2008. *Springer handbook of robotics*, volume 200. Springer.

Sünderhauf, N.; Brock, O.; Scheirer, W.; Hadsell, R.; Fox, D.; Leitner, J.; Upcroft, B.; Abbeel, P.; Burgard, W.; Milford, M.; et al. 2018. The limits and potentials of deep learning for robotics. *Int. J. Robot. Res.*, 37(4-5): 405–420.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press, second edition.

Tang, C.; Abbatematteo, B.; Hu, J.; Chandra, R.; Martín-Martín, R.; and Stone, P. 2024. Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes. *arXiv preprint arXiv:2408.03539*.

Wurman, P. R.; Barrett, S.; Kawamoto, K.; MacGlashan, J.; Subramanian, K.; Walsh, T. J.; Capobianco, R.; Devlic, A.; Eckert, F.; Fuchs, F.; et al. 2022. Outracing champion Gran Turismo drivers with deep reinforcement learning. *Nature*, 602(7896): 223–228.

Xiao, X.; Liu, B.; Warnell, G.; and Stone, P. 2022. Motion planning and control for mobile robot navigation using machine learning: a survey. *Auton. Robots*, 46(5): 569–597.

Yu, C.; Liu, J.; Nemati, S.; and Yin, G. 2021. Reinforcement learning in healthcare: A survey. *ACM Comput. Surv.*, 55(1): 1–36.