GCMRBench: Goal-Conditioned Multi-Robot Environments and Benchmarks for Advancing Offline Multi-Agent Reinforcement Learning

Extended Abstract

Chenxing Li
University of Tübingen
Tübingen, Germany
chenxing.li@student.uni-tuebingen.de

Chin-jui Chang
University of Tübingen
Tübingen, Germany
chin-jui.chang@student.uni-tuebingen.de

Yinlong Liu
City University of Macau
Macau, China
ylliu@cityu.edu.mo

Zijian Ma
Westlake University
Hangzhou, China
zijianma@westlake.edu.cn

Jan Seyler
Festo SE & Co. KG
Esslingen, Germany
jan.seyler@festo.com

Shahram Eivazi
University of Tübingen, Festo SE & Co. KG
Tübingen, Germany
shahram.eivazi@uni-tuebingen.de

ABSTRACT

Research in multi-agent reinforcement learning (MARL) has focused on developing algorithms to address challenges posed by agents' diverse goals, collaboration, and competition in complex environments. Extending these algorithms to Offline MARL (OMARL), together with the utilization of large-scale offline datasets, has increasingly been recognized as a promising approach toward safe, efficient, and rapid deployment in real-world scenarios. However, most existing studies train and evaluate OMARL in environments primarily designed for game-based scenarios. As a result, the potential of OMARL in domains such as robotics remains an open question. To bridge this gap, we introduce GCMRBench, a goal-conditioned multi-agent simulation environment tailored for dual-arm robotic tasks and the evaluation of offline multi-agent algorithms, thereby facilitating their deployment in practical robotic applications.

KEYWORDS

Offline Multi-Agent Reinforcement Learning; Benchmarks; Goal-Conditioned Dual-Robot Environments

ACM Reference Format:
Chenxing Li, Chin-jui Chang, Yinlong Liu, Zijian Ma, Jan Seyler, and Shahram Eivazi. 2026. GCMRBench: Goal-Conditioned Multi-Robot Environments and Benchmarks for Advancing Offline Multi-Agent Reinforcement Learning: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/HRLQ1652

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/HRLQ1652

1 INTRODUCTION

In recent years, a growing number of researchers have focused on multi-agent reinforcement learning (MARL) algorithms for solving complex cooperation tasks [1, 8, 9, 12, 17, 18, 21, 24]. One of the key areas of focus within this domain is OMARL, which leverages large-scale datasets to enable rapid and scalable deployment [5, 6, 10, 13, 16, 19, 21, 22, 29–31]. This paradigm minimizes the risks associated with online interactions, thereby improving the feasibility and safety of applying MARL methods in practical scenarios. However, the testing environments [14] for multi-agent algorithms are often limited to game-based or discrete settings [4, 20, 24, 25, 28], which pose challenges for evaluating their applicability in complex real-world tasks. Moreover, these environments are typically designed with fixed-step horizons that emphasize score maximization, which weakly reflects the deployment capability of OMARL algorithms in achieving success-oriented objectives under real-world conditions. To address these issues, we propose GCMRBench, a multi-agent environment designed for multi-robot collaboration that departs from traditional game-based settings and provides sufficiently realistic simulations of real-world robotic problems. The environment includes a variety of dual-arm robot tasks, establishing a task-oriented single and multi-agent framework, built upon PyBullet simulation [3] and Gymnasium API [27]. Our environment (Figure 1) is developed based on panda-gym [7] platform and comprises 22 dual-arm robot tasks, covering five categories: 1) ***Cooperation***, 2) ***Multi-Goal***, 3) ***Transition***, 4) ***Competition***, and 5) ***Hybrid*** tasks. Further, a total of 56 datasets, including both offline multi-agent and single-agent data, are provided to support comprehensive benchmarking with standard offline single and multi-agent reinforcement learning algorithms (MABC [26], IICQ [30], MABCQ [10], MACQL [11], InSPO [13], OMIGA [29]).

2 MULTI-ROBOT ENVIRONMENTS

In the design of our new environment, we incorporate relative observability for each agent to accommodate MARL. The evaluation protocol is task-oriented, with task success rate serving as

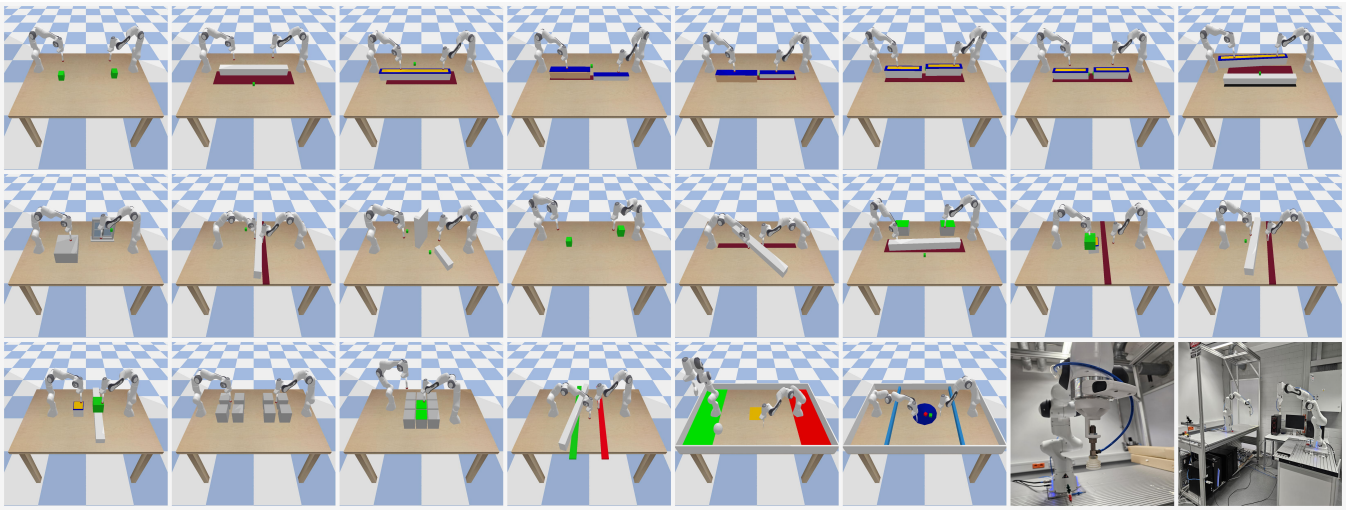

Figure 1: GCMRBench is dedicated to simulating industrial robotic tasks. The platform illustrates the 22 designed environments alongside their real-world robotic configurations. These environments are arranged from left to right and top to bottom, corresponding to the order of task descriptions presented later. The last images show the suction-enabled end-effector and the complete physical robot platform used for real-world experimental validation.

the primary performance metric. The framework supports both multi-agent and single-agent learning modes; in the multi-agent setting, each robotic arm is treated as an independent agent. In addition, human expert demonstrations are incorporated to guide data collection. This environment includes 11 ***Cooperation*** tasks (**Reach, Push, Lift, Insert, Connect, Insert(Goal-Changing), Connect(Goal-Changing), Stack, Window, Clamp, ReachAvoid**), 3 ***Multi-Goal*** tasks (ReachSeq, PushSeq, PushButton), 3 ***Transition*** tasks (**Handover, HandoverPush, AsynStack**), 4 ***Competition*** tasks (**Button, TicTacToe, CompetitionPush, Soccer**) and 1 ***Hybrid*** task (**Curling**), totaling 22 environments.

3 BENCHMARKS AND ANALYSIS

In this work, we collect a total of 56 datasets for offline learning (single and multi-settings). First, we use pretrained agents as experts to generate noisy rollouts at each step, which is a common approach for collecting expert data (*Pre-trained Data*). For medium-quality data, we select suboptimal pretrained agents (with success rates between 30% and 60%) and apply the same noisy rollout procedure. For poor-quality data (*Random Data*), actions are generated uniformly at random. We also mix the data with expert and poor data as *Mixed Data* to be used as medium data for training. The final type is *Guided Data* from human experts. We primarily use this platform to evaluate offline multi-agent algorithms, selecting the aforementioned algorithms. A subset of the results is presented here, as shown in the Table 1. Based on the overall performance, OMIGA and InSPO demonstrate significant advantages. They not only achieve strong performance on high-quality datasets but also exhibit superior data-exploitation capability on medium-quality datasets. We also observe training instability, characterized by oscillations in success rates and performance degradation after the peak success rate is achieved. In addition, we conduct benchmark evaluations for online learning and single-agent learning, and perform inference tests on real robotic systems [2, 15, 23]. The complete code, full paper contents, and all datasets will be open-sourced in the future.

Table 1: Performance comparison on selected environments. Results (%) are reported as mean ± sample standard deviation over five random runs after 20k training episodes. (G: Good, M: Medium, PT: Pretrained, Gui: Guided, Mix: Mixed.)

Environment	Type	MABC	OMIGA	InSPO
Lift	G-PT	1.00±0.00	1.00±0.00	0.99±0.01
	M-PT	0.42±0.07	0.45±0.06	0.48±0.08
Insert(Goal-Changing)	G-PT	0.97±0.02	0.96±0.02	0.95±0.02
	M-PT	0.34±0.04	0.37±0.07	0.60±0.06
ReachSeq	G-PT	0.90±0.04	0.88±0.06	0.65±0.13
	M-PT	0.34±0.16	0.45±0.16	0.49±0.23
CompetitionPush	G-PT	1.00±0.00	1.00±0.00	1.00±0.00
	M-PT	0.59±0.05	0.60±0.08	0.71±0.04
AsynStack	G-Gui	0.83±0.27	0.76±0.42	0.40±0.34
	M-Mix	0.18±0.13	0.12±0.10	0.12±0.16

4 CONCLUSION

This work introduces GCMRBench, the first task-oriented multi-agent simulation platform and dataset focusing on close to real-world robotic tasks. The known OMARL algorithms achieved Satisfactory results in offline settings compared to other benchmarks baselines, demonstrating the reliability of the simulation environment, datasets, and algorithms. However, the behavioral diversity of the datasets remains limited, the level of randomness in the environment may still be insufficient, and further research is needed to enhance algorithmic performance within this environment.

REFERENCES

[1] Nikita Chernyadev, Nicholas Backshall, Xiao Ma, Yunfan Lu, Younggyo Seo, and Stephen James. 2024. BiGym: A Demo-Driven Mobile Bi-Manual Manipulation Benchmark. *arXiv preprint arXiv:2407.07788* (2024).

[2] David Coleman, Ioan Sucan, Sachin Chitta, and Nikolaus Correll. 2014. Reducing the Barrier to Entry of Complex Robotic Software: a MoveIt! Case Study. arXiv:1404.3785 [cs.RO] https://arxiv.org/abs/1404.3785

[3] Erwin Coumans and Yunfei Bai. 2016–2021. PyBullet, a Python module for physics simulation for games, robotics and machine learning. http://pybullet.org.

[4] Rodrigo de Lazcano, Kallinteris Andreas, Jun Jet Tai, Seungjae Ryan Lee, and Jordan Terry. 2024. *Gymnasium Robotics*. http://github.com/Farama-Foundation/Gymnasium-Robotics

[5] Pu Feng, Junkang Liang, Size Wang, Xin Yu, Xin Ji, Yiting Chen, Kui Zhang, Rongye Shi, and Wenjun Wu. 2024. Hierarchical Consensus-Based Multi-Agent Reinforcement Learning for Multi-Robot Cooperation Tasks. arXiv:2407.08164 [cs.AI] https://arxiv.org/abs/2407.08164

[6] Claude Formanek, Asad Jeewa, Jonathan Shock, and Arnu Pretorius. 2023. Off-the-Grid MARL: Datasets with Baselines for Offline Multi-Agent Reinforcement Learning. arXiv:2302.00521 [cs.LG] https://arxiv.org/abs/2302.00521

[7] Quentin Gallouédec, Nicolas Cazin, Emmanuel Dellandréa, and Liming Chen. 2021. panda-gym: Open-Source Goal-Conditioned Environments for Robotic Learning. *4th Robot Learning Workshop: Self-Supervised and Lifelong Learning at NeurIPS* (2021).

[8] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. arXiv:1801.01290 [cs.LG] https://arxiv.org/abs/1801.01290

[9] Dom Huh and Prasant Mohapatra. 2024. Multi-agent Reinforcement Learning: A Comprehensive Survey. arXiv:2312.10256 [cs.MA] https://arxiv.org/abs/2312.10256

[10] Jiechuan Jiang and Zongqing Lu. 2023. Offline Decentralized Multi-Agent Reinforcement Learning. arXiv:2108.01832 [cs.LG] https://arxiv.org/abs/2108.01832

[11] Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. 2020. Conservative Q-Learning for Offline Reinforcement Learning. arXiv:2006.04779 [cs.LG] https://arxiv.org/abs/2006.04779

[12] Matthew Lai, Keegan Go, Zhibin Li, Torsten Kröger, Stefan Schaal, Kelsey Allen, and Jonathan Scholz. 2025. RoboBallet: Planning for multirobot reaching with graph neural networks and reinforcement learning. *Science Robotics* 10, 106 (2025), eads1204.

[13] Zongkai Liu, Qian Lin, Chao Yu, Xiawei Wu, Yile Liang, Donghui Li, and Xuetao Ding. 2024. Offline Multi-Agent Reinforcement Learning via In-Sample Sequential Policy Optimization. arXiv:2412.07639 [cs.AI] https://arxiv.org/abs/2412.07639

[14] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2020. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. arXiv:1706.02275 [cs.LG] https://arxiv.org/abs/1706.02275

[15] Christopher E. Mower, Theodoros Stouraitis, João Moura, Christian Rauch, Lei Yan, Nazanin Zamani Behabadi, Michael Gienger, Tom Vercauteren, Christos Bergeles, and Sethu Vijayakumar. 2022. ROS-PyBullet Interface: A Framework for Reliable Contact Simulation and Human-Robot Interaction. arXiv:2210.06887 [cs.RO] https://arxiv.org/abs/2210.06887

[16] Ling Pan, Longbo Huang, Tengyu Ma, and Huazhe Xu. 2022. Plan Better Amid Conservatism: Offline Multi-Agent Reinforcement Learning with Actor Rectification. arXiv:2111.11188 [cs.LG] https://arxiv.org/abs/2111.11188

[17] George Papadopoulos, Andreas Kontogiannis, Foteini Papadopoulou, Chaido Poulianou, Ioannis Koumentis, and George Vouros. 2025. An Extended Benchmarking of Multi-Agent Reinforcement Learning Algorithms in Complex Fully Cooperative Tasks. arXiv:2502.04773 [cs.LG] https://arxiv.org/abs/2502.04773

[18] Georgios Papoudakis, Filippos Christianos, Lukas Schäfer, and Stefano V. Albrecht. 2021. Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms in Cooperative Tasks. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS)*. http://arxiv.org/abs/2006.07869

[19] Seohong Park, Kevin Frans, Benjamin Eysenbach, and Sergey Levine. 2025. OGBench: Benchmarking Offline Goal-Conditioned RL. arXiv:2410.20092 [cs.LG] https://arxiv.org/abs/2410.20092

[20] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philiph H. S. Torr, Jakob Foerster, and Shimon Whiteson. 2019. The StarCraft Multi-Agent Challenge. *CoRR* abs/1902.04043 (2019).

[21] Eduardo Sebastian, Thai Duong, Nikolay Atanasov, Eduardo Montijano, and Carlos Sagues. 2025. Physics-Informed Multi-Agent Reinforcement Learning for Distributed Multi-Robot Problems. arXiv:2401.00212 [cs.RO] https://arxiv.org/abs/2401.00212

[22] Jianzhun Shao, Yun Qu, Chen Chen, Hongchang Zhang, and Xiangyang Ji. 2023. Counterfactual Conservative Q Learning for Offline Multi-agent Reinforcement Learning. arXiv:2309.12696 [cs.AI] https://arxiv.org/abs/2309.12696

[23] Stanford Artificial Intelligence Laboratory et al. [n.d.]. *Robotic Operating System*. https://www.ros.org

[24] J Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis S Santos, Clemens Dieffendahl, Caroline Horsch, Rodrigo Perez-Vicente, et al. 2021. Pettingzoo: Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021), 15032–15043.

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

[26] Faraz Torabi, Garrett Warnell, and Peter Stone. 2018. Behavioral Cloning from Observation. arXiv:1805.01954 [cs.AI] https://arxiv.org/abs/1805.01954

[27] Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, et al. 2024. Gymnasium: A Standard Interface for Reinforcement Learning Environments. *arXiv preprint arXiv:2407.17032* (2024).

[28] Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John Agapiou, Julian Schrittwieser, John Quan, Stephen Gaffney, Stig Petersen, Karen Simonyan, Tom Schaul, Hado van Hasselt, David Silver, Timothy Lillicrap, Kevin Calderone, Paul Keet, Anthony Brunasso, David Lawrence, Anders Ekermo, Jacob Repp, and Rodney Tsing. 2017. StarCraft II: A New Challenge for Reinforcement Learning. arXiv:1708.04782 [cs.LG] https://arxiv.org/abs/1708.04782

[29] Xiangsen Wang, Haoran Xu, Yinan Zheng, and Xianyuan Zhan. 2023. Offline Multi-Agent Reinforcement Learning with Implicit Global-to-Local Value Regularization. arXiv:2307.11620 [cs.LG] https://arxiv.org/abs/2307.11620

[30] Yiqin Yang, Xiaoteng Ma, Chenghao Li, Zewu Zheng, Qiyuan Zhang, Gao Huang, Jun Yang, and Qianchuan Zhao. 2021. Believe What You See: Implicit Constraint Approach for Offline Multi-Agent Reinforcement Learning. arXiv:2106.03400 [cs.AI] https://arxiv.org/abs/2106.03400

[31] Zhengbang Zhu, Minghuan Liu, Liyuan Mao, Bingyi Kang, Minkai Xu, Yong Yu, Stefano Ermon, and Weinan Zhang. 2025. MADiff: Offline Multi-agent Learning with Diffusion Models. arXiv:2305.17330 [cs.AI] https://arxiv.org/abs/2305.17330