Mixed Human-UAV Reinforcement Learning: Literature Review and Open Challenges

Nicolo' Brandizzi*[a], Damiano Brunori*[b], Francesco Frattolillo*[c], Alessandro Trapasso*[d] and Luca Iocchi[e]

Department of Computer, Automation and Management Engineering, Sapienza University of Rome, via Ariosto 25 Roma 00185, Italy

Keywords: UAVs, Reinforcement Learning, Human-in-the-Loop, Human-UAV.

Abstract: Unmanned Aerial Vehicles (UAVs) are becoming a popular solution for a plethora of tasks, ranging from supporting and extending communication to monitoring and exploring areas of interest. At the same time, Reinforcement Learning (RL) has become an excellent candidate technique to face complex scenarios where a model of the environment is not always available. Nevertheless, fully autonomous applications can have some drawbacks under certain unpredictable circumstances. Thus an active human element could facilitate handling such scenarios. All these things considered, and after an in-depth literature analysis, we focused on Mixed Human-UAV reinforcement learning applications that would benefit from introducing the human-in-the-loop component by pointing out their strengths, weakness, and new challenges.

1 INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are increasingly becoming a key technology to solve a wide range of problems due to their high mobility in three-dimensional space, easy deployment, and relatively low cost. Some examples include UAVs used as aerial stations to provide services after a natural disaster or for network access in remote areas, for emergency communications and rescue support, when performing surveillance tasks in risky and dangerous areas, and also for cargo and goods delivery to supply food and medical goods and in the agricultural field to monitor and facilitate farming activities.

Autonomous UAV. As UAV application fields are increasing in recent years, a general research trend is aiming at developing autonomous UAVs able to make their own real-time decisions, thus ensuring a highly responsive autonomous system in dynamic and even unknown possible scenarios. Indeed, a high degree of automation could lead UAVs to (co)operate without the need for constant device-to-device communication and to avoid complex and slow human remote-controlled maneuvers.

An excellent approach that has proven extremely effective in various complex problems is *Reinforcement Learning* (RL). In recent years, a number of problems have been successfully solved by Deep Reinforcement Learning (DRL), which is an integration of deep neural networks in RL. In fact, it has been shown that agents trained through DRL were able to obtain super-human performance in extremely complex games such as the classic Atari games, Go, Dota 2, and complex robotics problems.

Multi-UAV Reinforcement Learning. UAVs are designed to operate in shared and dynamic airspace, hence cooperation is desirable and necessary.

We are thus interested in Multi-UAV systems, i.e., teams or fleets of UAVs (Ali et al., 2021; Duflo et al., 2020) cooperating to achieve the desired goal. By leveraging a team of UAVs, tasks can be executed faster and more efficiently than by using a single UAV. UAVs also need to achieve their goal while avoiding conflicts and collisions with other UAVs and objects in the environment, since they are deployed in shared airspace and safety (including ground safety) is of utmost importance.

[a] https://orcid.org/0000-0002-3191-6623
[b] https://orcid.org/0000-0002-0384-3498
[c] https://orcid.org/0000-0002-2040-3355
[d] https://orcid.org/0000-0001-5431-6607
[e] https://orcid.org/0000-0001-9057-8946
* Authors are alphabetically ordered

Brandizzi, N., Brunori, D., Frattolillo, F., Trapasso, A. and Iocchi, L.
Mixed Human-UAV Reinforcement Learning: Literature Review and Open Challenges.
DOI: 10.5220/0011955100003622
In *Proceedings of the 1st International Conference on Cognitive Aircraft Systems (ICCAS 2022)*, pages 38-42
ISBN: 978-989-758-657-6

Most of the relevant RL and DRL literature targets single-agent systems. However, in this paper, we focus on the application of RL in a multi-agent system (MAS), known as Multi-Agent Reinforcement Learning (MARL) (Jafari et al., 2020), and more specifically on Multi-UAV RL techniques.

There are a number of reinforcement learning techniques applied to the multi-agent scenario. For example, independent Learners (Tan, 1993) are used for UAV trajectory planning and time resource allocation in a multiple UAV-enabled network (Tang et al., 2020). A slightly different problem such as trajectory planning for Multi-UAV assisted Mobile edge-computing (EG), (Wang et al., 2021), can be faced with Multi-Agent Deep Deterministic Policy Gradient (MADDPG) (Lowe et al., 2017) algorithm. Moreover, an extension of the famous Deep Q-learning algorithm (Mnih et al., 2013) to the Multi-agent case is applied for autonomous forest fire fighting (Haksar and Schwager, 2018): this approach is based on an exploratory strategy led by an optimal heuristic function. Other scenarios may arise where (D)RL can be efficiently deployed, e.g., leader-follower systems (Hung and Givigi, 2017), long-term communication coverage (Liu et al., 2020a), Mobile Crowd Sensing (MCS) (Liu et al., 2019).

2 OVERVIEW OF Multi-UAV RL METHODS

In this section, we present a brief review of the literature on Reinforcement Learning applied to Multi-UAV systems. Initially, we investigated major conferences such as AAMAS (International Conference on Autonomous Agents and Multiagent Systems), ECML (European conference on machine learning), ICML (The International Conference on Machine Learning), NeurIPS (Neural Information Processing Systems), but without relevant matches. Thus we moved to a keyword search on scientific databases such as *IEEE Xplore*, *ScienceDirect*, *MDPI* and *Springer*. Our interest was focused on the following keywords: *UAVs*, *Multi-UAVs*, *Reinforcement learning*, *Drones*, *Multi-agents* and a combination of them. As a consequence of our analysis and as shown in Figure 1, this topic is recent and rapidly growing.

Additionally, the European Drone Outlook Study (Undertaking, 2017) forecasts an increase in the drone marketplace of *EUR 15 billion annually* by 2050 and they estimated at least EUR 200 million in additional funding in Research & Development within 5 to 10 years from the study. Furthermore, they expect around 7 million consumer leisure drones to operate

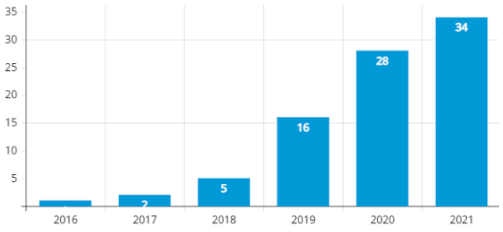

Figure 1: Number of publications per year (2016 to 2021) given the research keywords.

across Europe and around 400 thousand for government and commercial missions in 2050. The above-mentioned data reveals the interest in this emerging sector and therefore the need to find new solutions for their uses. (Undertaking, 2017) also points out the sectors which will benefit the most as being *agriculture, energy, E-commerce, delivery* and *transport*.

By analyzing the literature, our aim is to highlight the most popular research directions and group-specific problems into macro classes such as Coverage, Trajectory generation, Computation Offloading and Communication. In this regard, we noticed a lack of work including human supervision on autonomous UAVs in case of unexpected events. Although machine learning (ML) and in particular RL are excellent tools capable of automating the agents' behavior, humans still play an important role in supervising UAVs.

Mixed Human-Robot Teams. Fully autonomous agents are prone to errors, as they cannot cope with the amount of knowledge required to successfully perform a task in a real-world environment. Several studies investigated the level of autonomy (LOA) adopted by agents in mixed human-robot settings (Wu et al., 2018). The autonomy is most often analyzed in terms of a dynamic framework in which the agents learn to adapt their LOA to the task at hand, e.g., mixed-initiative interaction settings (Allen et al., 1999), adaptive autonomy (Suzanne Barber et al., 2000) and collaborative control (Fong et al., 2001). MAS with autonomous agents are sensitive to a range of issues related to their performance for computer vision tasks, such as object detection, imperfect knowledge of the system, and lastly ethical issues resulting from non-explainable AI decisions.

During our literature research, we found only one work focusing on mixed human-robot teams for multi-UAVs systems (WANG et al., 2020). In the latter, the authors build a transparent human-UAV framework where they study the behavior of heterogeneous UAVs equipped with different capabilities. This framework is based on the requirements of *Observability*, *Predictability*, and *Directionality* (OPD),

which must be met for each agent on the team to understand the intentions and behaviors of the other members, contrary to previous approaches which focused on improving the level of autonomy (LOA) of team members. In this work, agents can mitigate this problem by being aware of the state and actions of other agents. An optimal policy for the leader-follower problem is learned through deep reinforcement learning (DRL), while a path planner is used to take into account the presence of enemies threats in dynamic environments. These policies help humans by providing hints or warnings if the Human user, a.k.a. Manned Aerial Vehicle (MAV), approaches dangerous areas. MAV behaviors are understandable and controllable, allowing humans to observe, predict and direct them.

3 OPEN SCIENTIFIC CHALLENGES & APPLICATIONS

In our analysis of the current state of the art on reinforcement learning for Multi-UAV systems, we found a lack of work considering the presence of one or more human agents within the environment. When considering mixed human-robot interactions, the issue of interpretability becomes crucial for a human operator to understand and guide autonomous UAVs. This problem is well addressed in the field of explainable-AI (XAI), whose aim is to stir from the classical black-box approach in favor of a more user-friendly system. We advocate that mixed human-robot teams are fundamental for solving complex tasks in a shared aerial environment and the related research activities should proceed in this direction. In the following section, we present some promising research lines and industrial applications that can benefit from more in-depth studies in this field. All the presented applications can be built using a MARL framework and focus on the supervision and/or collaboration applied by a human operator on multiple UAVs for a range of different scenarios.

3.1 Challenges

Some research directions are here considered to investigate the feasibility and safety of a human-UAV hybrid RL system.

Explainable UAVs Behaviour. To supervise autonomous UAVs, humans need to understand the decision-making process of AI models. For this reason, we want to avoid the classic black-box approach in favor of a more understandable method (WANG et al., 2020). UAVs can hardly have full autonomy to solve difficult tasks regardless of the human intervention, and this becomes even more evident when ethical reasons are involved in decisional processes: human-in-the-loop is still crucial in this case. Therefore, UAVs need to cooperate with human users according to protocols that should be as clear as possible in order to avoid misunderstanding issues between drones and humans' behaviors. As a result, research should focus on the balance between the autonomy level of UAVs and human control: in this regard, it is essential to represent the behaviors learned by the drones through DRL in some formalism that is clear and understandable to a human operator.

Knowledge Representation. To the best of our knowledge, there is a noticeable lack of works modeling the presence of a human in the context of RL, which mainly relies on the Markov Decision Process (MDP) formalism for its problems description. Formally an MDP is defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$, where the elements are respectively the state and the action spaces of agents and the transition and reward functions of an environment. Extending an MDP to take into account the human-in-the-loop factor could be a very difficult challenge. Several works such as (Gateau et al., 2016) involving both humans and UAVs, model the former through a Mixed Observability MDP (MOMDP). MOMDP is an extension of MDP with partial observations and stochastic policies. In the literature the use of MOMDP is usually associated with Planning algorithms, while their application in Mixed Human-UAV are still unexplored. Furthermore, we need also to consider that standard agents in a generic RL problem require a number of learning interactions with the environment much higher than their human counterparts. On the other hand, a non-human agent provided with high computational capability can predict some specific information that may easily be neglected by a human agent.

Conflict Resolution. When dealing with UAVs, Air traffic monitoring (ATM) becomes an unmanned ATM, i.e. UTM, and in this latter case obviously more particular and additional safety constraints are required with respect to the ATM case. Thus, when an emergency event arises in a given UTM scenario with autonomous UAVs, a responsive human intervention can be necessary and resolutive. Autonomous to human control switch is desirable when a severe and not delayable fault is happening and a manually man-

aged system is extremely required. In these particular conditions, the human user should directly control the faulty UAV in order to allow manual emergency maneuvers. Acting in this way, motions or even landing operations can be performed in complete safety while the other UAV swarm members fly along new paths which are recomputed online by applying (D)RL deconfliction algorithms (Isufaj et al., 2021). This approach can be applied to any kind of conflict-risky or emergency scenario which can occur at any flight phase, namely either during the ongoing flight or in takeoff or landing phases.

3.2 Applications

Among the plethora of available applications for UAV systems, we want to focus attention on some of those who could benefit from adding a human-in-the-loop element.

UAV Relay Networks. Data gathering, edge-computing, and information flow are crucial in modern scenarios and present a number of challenges tied to the hostile nature of the environment. Centralized networks control is the preferred structure in state-of-the-art systems, however, it does not guarantee an adequate level of robustness when dealing with disturbing phenomena such as interference and communication jamming (Wang et al., 2020). On the other hand, a decentralized Multi-UAV system can be deployed in order to guarantee a sufficient amount of autonomous behavior. The system can leverage the reinforcement learning framework's ability to focus on multiple goals concurrently such as maximizing the physical connectivity among UAVs and the area of coverage. Finally, the system can be supervised and controlled by a human agent on the field only by transmitting high-level actions, e.g. *move forward*, instead of coordinating the whole UAVs team.

Improved UAVs Surveillance. A possible application of Multi-UAVs systems and DRL could be persistent coverage and surveillance (Liu et al., 2020b). In this context, the UAVs must monitor a certain area through the use of sensors such as onboard RGB cameras, trying to organize themselves in order to offer adequate coverage and therefore also trying to minimize the overlap of the monitored sub-areas. In a standard application, individual UAVs can send notifications in case of suspicious events in a given zone and proceed again with their workflow. In an alternative version of the same application but with the human-in-the-loop component, the human operator could be asked to take control of the UAV through teleoperation, following the notification. In this case, the human operator would be able to control the UAV and therefore follow and accurately track the event. In this way, it is also possible to reduce the amount of data exchanged, as the UAV is only delegated to send notifications while the real-time streaming of the monitored environment can be activated following the human intervention.

4 CONCLUSION

In this work, we describe various UAVs applications and highlight some evidence showing that the UAV sector is constantly growing in the last few years. In particular, we focus on a literature analysis related to RL approaches applied to Multi-UAV systems. As a result, we identify a shortage of works concerning the human role as possible collaborator or supervisor in complex mixed scenarios involving unexpected and not easily predictable events: most of the time these particular conditions are challenging and problematic to be managed by the UAVs autonomously. Finally, we propose some alternative solutions to existing problems, trying to indicate how the presence of a human operator can be crucial and have a positive contribution in finding a valid and possibly optimal solution.

ACKNOWLEDGEMENTS

This paper has been partially supported by BUBBLES Project. BUBBLES project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation program under grant agreement No 893206. Research has been partially supported also by the ERC Advanced Grant WhiteMech (No. 834228) and by the EU ICT-48 2020 project TAILOR (No. 952215).

REFERENCES

Ali, Z. A., Han, Z., and Masood, R. J. (2021). Collective motion and self-organization of a swarm of uavs: A cluster-based architecture. *Sensors*, 21(11):1–19.

Allen, J. E., Guinn, C. I., and Horvtz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications*, 14(5):14–23.

Duflo, G., Danoy, G., Talbi, E. G., and Bouvry, P. (2020). Automating the Design of Efficient Distributed Behaviours for a Swarm of UAVs. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, pages 489–496.

Fong, T., Thorpe, C., and Baur, C. (2001). *Collaborative control: A robot-centric model for vehicle teleoperation*, volume 1. Carnegie Mellon University, The Robotics Institute Pittsburgh.

Gateau, T., Chanel, C. P. C., Le, M.-H., and Dehais, F. (2016). Considering human's non-deterministic behavior and his availability state when designing a collaborative human-robots system. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4391–4397.

Haksar, R. N. and Schwager, M. (2018). Distributed Deep Reinforcement Learning for Fighting Forest Fires with a Network of Aerial Robots. *IEEE International Conference on Intelligent Robots and Systems*, pages 1067–1074.

Hung, S. M. and Givigi, S. N. (2017). A Q-Learning Approach to Flocking with UAVs in a Stochastic Environment. *IEEE Transactions on Cybernetics*, 47(1):186–197.

Isufaj, R., Omeri, M., and Piera, M. A. (2021). Multi-uav conflict resolution with graph convolutional reinforcement learning. *CoRR*, abs/2111.14598.

Jafari, M., Xu, H., and Carrillo, L. R. G. (2020). A biologically-inspired reinforcement learning based intelligent distributed flocking control for Multi-Agent Systems in presence of uncertain system and dynamic environment. *IFAC Journal of Systems and Control*, 13:100096.

Liu, C. H., Chen, Z., and Zhan, Y. (2019). Energy-Efficient Distributed Mobile Crowd Sensing: A Deep Learning Approach. *IEEE Journal on Selected Areas in Communications*, 37(6):1262–1276.

Liu, C. H., Ma, X., Gao, X., and Tang, J. (2020a). Distributed Energy-Efficient Multi-UAV Navigation for Long-Term Communication Coverage by Deep Reinforcement Learning. *IEEE Transactions on Mobile Computing*, 19(6):1274–1285.

Liu, Y., Liu, H., Tian, Y., and Sun, C. (2020b). Reinforcement learning based two-level control framework of UAV swarm for cooperative persistent surveillance in an unknown urban area. *Aerospace Science and Technology*, 98:105671.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., and Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *CoRR*, abs/1706.02275.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing atari with deep reinforcement learning.

Suzanne Barber, K., Goel, A., and Martin, C. E. (2000). Dynamic adaptive autonomy in multi-agent systems. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(2):129–147.

Tan, M. (1993). Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents. *Mach. Learn. Proc. 1993*, pages 330–337.

Tang, J., Song, J., Ou, J., Luo, J., Zhang, X., and Wong, K. K. (2020). Minimum Throughput Maximization for Multi-UAV Enabled WPCN: A Deep Reinforcement Learning Method. *IEEE Access*, 8:9124–9132.

Undertaking, S. E. S. A. R. . J. (2017). *European drones outlook study : unlocking the value for Europe*. Publications Office.

WANG, C., WU, L., YAN, C., WANG, Z., LONG, H., and YU, C. (2020). Coactive design of explainable agent-based task planning and deep reinforcement learning for human-UAVs teamwork. *Chinese Journal of Aeronautics*, 33(11):2930–2945.

Wang, L., Wang, K., Pan, C., Xu, W., Aslam, N., and Hanzo, L. (2021). Multi-Agent Deep Reinforcement Learning-Based Trajectory Planning for Multi-UAV Assisted Mobile Edge Computing. *IEEE Transactions on Cognitive Communications and Networking*, 7(1):73–84.

Wang, W., Lu, X., Liu, S., Xiao, L., and Yang, B. (2020). Energy Efficient Relay in UAV Networks Against Jamming: A Reinforcement Learning Based Approach. *IEEE Vehicular Technology Conference*, 2020-May:0–4.

Wu, X., Wang, C., Niu, Y., Hu, X., and Fan, C. (2018). Adaptive human-in-the-loop multi-target recognition improved by learning. *International Journal of Advanced Robotic Systems*, 15(3):1729881418774222.