Autonomous Agent for Beyond Visual Range Air Combat: A Deep Reinforcement Learning Approach

Joao P. A. Dantas
dantasjpad@fab.mil.br
Institute for Advanced Studies
Sao Jose dos Campos, Brazil

Marcos R. O. A. Maximo
mmaximo@ita.br
Aeronautics Institute of Technology
Sao Jose dos Campos, Brazil

Takashi Yoneyama
takashi@ita.br
Aeronautics Institute of Technology
Sao Jose dos Campos, Brazil

ABSTRACT

This work contributes to developing an agent based on deep reinforcement learning capable of acting in a beyond visual range (BVR) air combat simulation environment. The paper presents an overview of building an agent representing a high-performance fighter aircraft that can learn and improve its role in BVR combat over time based on rewards calculated using operational metrics. Also, through self-play experiments, it expects to generate new air combat tactics never seen before. Finally, we hope to examine a real pilot's ability, using virtual simulation, to interact in the same environment with the trained agent and compare their performances. This research will contribute to the air combat training context by developing agents that can interact with real pilots to improve their performances in air defense missions.

CCS CONCEPTS

• **Computing methodologies** → **Simulation support systems**;
• **Applied computing** → **Military**.

1 INTRODUCTION

Air combat is a complex and dynamic scenario involving highly skilled pilots making split-second decisions to outmaneuver their opponents [7]. Beyond visual range (BVR) air combat, in particular, presents unique challenges and opportunities, as it involves engagements taking place at ranges beyond the pilot's ability to see the enemy aircraft [10]. While modern air combat may still be within visual range (WVR), the combat usually begins in BVR. This phase is frequently the most critical as it can provide advantages and drawbacks for succeeding phases [2]. The fundamental difficulty for pilots in the fight is maneuver planning, which reflects both sides' tactical decision-making capacity and determines success or failure [11]. Constructive computer simulations can mimic the most diverse BVR combat situations to investigate the effects of

new combat tactics, sensors, and armaments [6]. One of the main challenges of constructive BVR combat simulation is to simulate the complex behaviors of a pilot in all phases of combat. A pilot can perform decision-making processes such as adapting to new combat situations and conducting collective tactics with pilots from other aircraft to employ an engagement [3], or launching a missile at the proper instant [4].

The presented work proposes a model to create an autonomous agent utilizing deep reinforcement learning (DRL) to operate in a BVR air combat simulation environment. By learning from operational metrics, the agent will represent a high-performance fighter aircraft that can enhance its capabilities over time. Furthermore, we expected the generation of novel air combat tactics through self-play experiments. The ultimate goal is to allow real pilots to engage with the trained agent in the same environment using virtual simulation and compare their performances. The framework's primary objectives are to design a BVR agent that can learn all the tactics associated with this air combat mode, enhance its performance through self-play techniques, and outperform a real pilot in the context of BVR air combat.

2 RELATED WORK

Several recent studies have explored DRL algorithms for autonomous decision-making in BVR air combat scenarios for different applications: generation of air combat tactics [15], maneuver planning [8, 11, 18, 19], and multi-UAV cooperative decision-making methods [12, 13]. These works demonstrate the potential of DRL-based approaches for decision-making in BVR air combat scenarios. While these approaches have shown promising results, there is still much work to do in developing more robust and sufficient algorithms and evaluating the feasibility of these methods in real-world applications. In contrast to these studies, our work focuses on applying DRL techniques to BVR air combat using a high-fidelity simulation environment. Besides, to the best of the authors' knowledge, no study has yet explored the use of DRL, combined with self-play techniques, to train a high-performance agent that can interact with a real pilot in the same simulation environment.

3 PROPOSED MODEL

The Rainbow algorithm [9] is a state-of-the-art DRL technique that combines several improvements to the Deep Q-Network (DQN) [14] algorithm to achieve better results. We will use this algorithm to train our autonomous BVR air combat agent. The Rainbow algorithm includes some improvements over the DQN algorithm, including prioritized experience replay, distributional reinforcement learning, and dueling networks. Prioritized experience replay [16] allows the agent to prioritize specific transitions in the replay buffer

based on their importance for learning. At the same time, distributional reinforcement learning [1] estimates the distribution of the discounted return instead of just the expected value. Finally, dueling networks [17] separate the estimation of the state value and the advantage function to improve the learning process.

We will use the Aerospace Simulation Environment (*Ambiente de Simulação Aeroespacial – ASA* in Portuguese) as the simulation platform to train and evaluate our agent. ASA is a custom-made, high-fidelity, object-oriented simulation framework developed mainly in C++ that enables the modeling and simulation of military operational scenarios to support the development of tactics and procedures in the aerospace context for the Brazilian Air Force [5].

To manage the agent's actions $a(t)$, we will consider the following possible tactics: Combat Air Patrol (CAP), Commit, Abort, Break, Fire, and Support. CAP involves flying around a particular location in a specific pattern; Commit means to engage a detected target; Abort is a tactic of moving away from a threat; Break is a sudden defensive maneuver when the radar detects a missile; Fire is launching a missile at a target; and Support is the tactic to aid the missile's guidance until its seeker become activate [3, 4]. The agent's state $s(t)$ depends on the independent motion variables of the agent, such as position $[p_x(t), p_y(t), p_z(t)]$, velocity $[v_x(t), v_y(t), v_z(t)]$, and orientation regarding roll $\phi(t)$, pitch $\theta(t)$, and yaw $\psi(t)$. Besides, the state includes comparative factors between the agent and the nearest detected target, such as relative distance $\Delta d(t)$, relative speed $\Delta v(t)$, and relative angle $\Delta \alpha(t)$. Finally, the last part of the agent's state is composed of the remaining fuel $f(t)$, the remaining missiles $m(t)$, the health condition $h(t)$, and the status of the agent's sensors $s_s(t)$. During training, the Rainbow algorithm will update the neural network based on the rewards received by the agent $r(t)$, calculated by the Defensive Counter Air (DCA) Index [3], a probability of success for BVR combat on DCA missions whose objective is to establish a CAP. Figure 1 illustrates the agent's interaction with the environment.
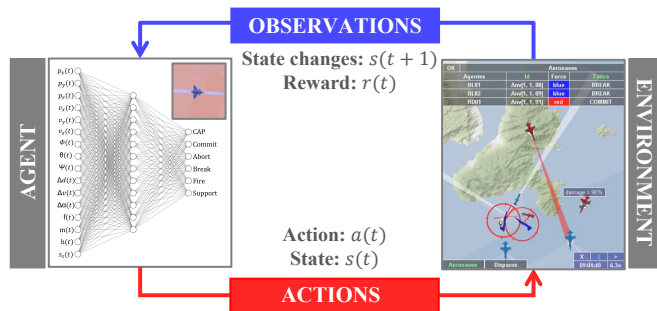

Figure 1: Agent-environment interaction: a_t, s_t, **and** r_t **denote action, state, and reward at time step** t, **with** s_{t+1} **given by the environment for the next iteration.**

4 CONCLUSIONS

This research aims to improve air combat training by developing unmanned combat aerial vehicles (UCAVs) to interact with pilots and enhance fighter performance in air defense missions. We will

release the source code for the general architecture to encourage the development of diverse applications using the same simulation platform. The technology developed in this work has the potential to serve as a simulation-as-a-service (SimaaS) tool to meet various simulation needs in the defense and aerospace industries.

REFERENCES

[1] Marc G Bellemare, Will Dabney, and Rémi Munos. 2017. A distributional perspective on reinforcement learning. In *International Conference on Machine Learning*. PMLR, 449–458.

[2] Joao P. A. Dantas. 2018. *Apoio à Decisão para o Combate Aéreo Além do Alcance Visual: Uma Abordagem por Redes Neurais Artificiais.* Master's Thesis. Instituto Tecnológico de Aeronáutica, São José dos Campos, SP, Brazil.

[3] Joao P. A. Dantas, Andre N. Costa, Diego Geraldo, Marcos R. O. A. Maximo, and Takashi Yoneyama. 2021. Engagement Decision Support for Beyond Visual Range Air Combat. In *Proceedings of the 2021 Latin American Robotics Symposium, 2021 Brazilian Symposium on Robotics, and 2021 Workshop on Robotics in Education*. October 11th–15th, 96–101.

[4] Joao P. A. Dantas, Andre N. Costa, Diego Geraldo, Marcos R. O. A. Maximo, and Takashi Yoneyama. 2021. Weapon Engagement Zone Maximum Launch Range Estimation Using a Deep Neural Network. In *Intelligent Systems*, André Britto and Karina Valdivia Delgado (Eds.). Springer, Cham, 193–207.

[5] Joao P. A. Dantas, Andre N. Costa, Vitor C. F. Gomes, Andre R. Kuroswiski, Felipe L. L. Medeiros, and Diego Geraldo. 2022. ASA: A Simulation Environment for Evaluating Military Operational Scenarios. In *Proceedings of the 20th International Conference on Scientific Computing*. 25th–28th, Las Vegas, NV, USA.

[6] Joao P. A. Dantas, Andre N. Costa, Felipe L. L. Medeiros, Diego Geraldo, Marcos R. O. A. Maximo, and Takashi Yoneyama. 2022. Supervised Machine Learning for Effective Missile Launch Based on Beyond Visual Range Air Combat Simulations. In *Proceedings of the Winter Simulation Conference* (Singapore) (*WSC '22*).

[7] Joao P. A. Dantas, Marcos R. O. A. Maximo, Andre N. Costa, Diego Geraldo, and Takashi Yoneyama. 2022. Machine Learning to Improve Situational Awareness in Beyond Visual Range Air Combat. *IEEE Latin America Transactions* 20, 8 (2022). https://latamt.ieeer9.org/index.php/transactions/article/view/6530

[8] Zihao Fan, Yang Xu, Yuhang Kang, and Delin Luo. 2022. Air Combat Maneuver Decision Method Based on A3C Deep Reinforcement Learning. *Machines* 10, 11 (2022), 1033.

[9] Matteo Hessel, Joseph Modayil, Hado van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Gheshlaghi Azar, and David Silver. 2018. Rainbow: Combining Improvements in Deep Reinforcement Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32. 3150–3157.

[10] Lt Col Patrick Higby and USAF Col. 2005. Promise and reality: Beyond visual range (BVR) air-to-air combat. *Air War College (AWC) Electives Program: Air Power Theory, Doctrine, and Strategy: 1945–Present* 30 (2005).

[11] Dongyuan Hu, Rennong Yang, Jialiang Zuo, Ze Zhang, Jun Wu, and Ying Wang. 2021. Application of deep reinforcement learning in maneuver planning of beyond-visual-range air combat. *IEEE Access* 9 (2021), 32282–32297.

[12] Jinwen Hu, Luhe Wang, Tianmi Hu, Chubing Guo, and Yanxiong Wang. 2022. Autonomous maneuver decision making of dual-UAV cooperative air combat based on deep reinforcement learning. *Electronics* 11, 3 (2022), 467.

[13] Xiaoxiong Liu, Yi Yin, Yuzhan Su, and Ruichen Ming. 2022. A Multi-UCAV Cooperative Decision-Making Method Based on an MAPPO Algorithm for Beyond-Visual-Range Air Combat. *Aerospace* 9, 10 (2022), 563.

[14] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529–533.

[15] Haiyin Piao, Zhixiao Sun, Guanglei Meng, Hechang Chen, Bohao Qu, Kuijun Lang, Yang Sun, Shengqi Yang, and Xuanqi Peng. 2020. Beyond-visual-range air combat tactics auto-generation by reinforcement learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[16] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. 2015. Prioritized Experience Replay. *arXiv preprint arXiv:1511.05952* (2015).

[17] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado van Hasselt, Marc Lanctot, and Nando de Freitas. 2015. Dueling network architectures for deep reinforcement learning. In *International Conference on Machine Learning*. PMLR, 1995–2003.

[18] Hongpeng Zhang, Yujie Wei, Huan Zhou, and Changqiang Huang. 2022. Maneuver Decision-Making for Autonomous Air Combat Based on FRE-PPO. *Applied Sciences* 12, 20 (2022), 10230.

[19] Hongpeng Zhang, Huan Zhou, Yujie Wei, and Changqiang Huang. 2022. Autonomous maneuver decision-making method based on reinforcement learning and Monte Carlo tree search. *Frontiers in Neurorobotics* (2022).