

Penelitian ini bertujuan merancang dan mengevaluasi sistem navigasi untuk robot tumbuh lunak 3D pada lingkungan dengan rintangan statis, menggunakan tiga algoritma *Deep Reinforcement Learning* (DRL), yaitu *Proximal Policy Optimization* (PPO), *Deep Q-Network* (DQN), dan *Advantage Actor-Critic* (A2C). Implementasi dilakukan sepenuhnya berbasis Python dengan *framework* Stable-Baselines3, serta lingkungan simulasi yang dibangun di OpenAI Gym. Lingkungan 3D mencakup target statis yang posisinya berubah pada setiap percobaan dan rintangan yang tetap, untuk menguji kemampuan robot dalam mengatur lintasan dan menghindari tabrakan. *Action space* didefinisikan secara diskrit (maju, belok kiri, belok kanan, diam), sedangkan *observation space* mencakup posisi robot, posisi target, dan parameter navigasi lainnya. Fungsi *reward* berbentuk logaritmik terhadap jarak ke target, dilengkapi penalti untuk tabrakan, sehingga mendorong strategi navigasi yang efisien.

Hasil pengujian menunjukkan bahwa PPO secara konsisten menghasilkan lintasan yang lebih halus dan akurat, dengan nilai MSE akhir yang lebih rendah dibandingkan DQN dan A2C. Nilai MSE akhir PPO berkisar antara 0,010 hingga 0,075, menunjukkan kemampuan PPO dalam mengurangi kesalahan menuju target dengan stabil, baik dalam menghindari rintangan maupun mencapai target. Sebaliknya, DQN cenderung melakukan koreksi agresif yang menghasilkan nilai MSE akhir yang lebih tinggi, berkisar antara 0,412 hingga 2,268, dan menghasilkan lintasan yang kurang stabil. A2C menunjukkan hasil yang lebih baik dibandingkan DQN, dengan nilai MSE akhir yang berkisar antara 0,255 hingga 1,349, meskipun ada variabilitas dalam kecepatan adaptasi. Percobaan dengan waktu penyelesaian lebih cepat, seperti *Trial 7* (0,50 menit), menunjukkan hasil lebih optimal, dengan *clearance* yang cepat mencapai stabilitas di sekitar nilai 1,0.

Temuan ini menegaskan bahwa PPO lebih unggul dalam hal stabilitas dan efisiensi terhadap variasi posisi target dan rintangan, menjadikannya pendekatan yang lebih andal untuk navigasi robot tumbuh lunak 3D, dengan aplikasi potensial pada eksplorasi ruang sempit, inspeksi saluran tertutup, operasi pencarian dan penyelamatan di area reruntuhan, serta aplikasi medis minimal invasif. PPO terbukti lebih stabil dan efisien dalam mencapai target dibandingkan dengan DQN dan A2C, yang lebih terpengaruh oleh dinamika lingkungan yang kompleks.

Kata kunci—navigasi, robot tumbuh lunak 3D, PPO, DQN, A2C, *observation space*, *action space*.

ABSTRACT

This study aims to design and evaluate a navigation system for 3D soft growing robots in a static obstacle environment using three Deep Reinforcement Learning (DRL) algorithms: Proximal Policy Optimization (PPO), Deep Q-Network (DQN), and Advantage Actor-Critic (A2C). The implementation is entirely based on Python with the Stable-Baselines3 framework, and a simulation environment built in OpenAI Gym. The 3D environment includes static targets whose positions change with each trial and fixed obstacles to test the robot's ability to plan its path and avoid collisions. The action space is defined discretely (move forward, turn left, turn right, stay), while the observation space includes the robot's position, target position, and other supporting navigation parameters. The reward function is logarithmic concerning the distance to the target, with a penalty for collisions, encouraging efficient navigation strategies.

The test results show that PPO consistently produces smoother and more accurate paths, with final MSE values lower than DQN and A2C. The final MSE values for PPO range from 0.010 to 0.075, indicating that PPO effectively reduces the error towards the target while maintaining stability in avoiding obstacles and reaching the target. In contrast, DQN tends to make more aggressive corrections, resulting in higher final MSE values, ranging from 0.412 to 2.268, producing less stable trajectories. A2C showed better results than DQN, with final MSE values ranging from 0.255 to 1.349, though there was variability in the adaptation speed. Trials with quicker completion times, such as Trial 7 (0.50 minutes), showed more optimal results, with the clearance quickly stabilizing around 1.0.

These findings confirm that PPO outperforms DQN and A2C in terms of stability and efficiency to variations in target positions and obstacles, making it a more reliable approach for 3D soft growing robot navigation, with potential applications in narrow space exploration, closed-channel inspection, search and rescue operations in rubble, and minimally invasive medical applications. PPO proves to be more stable and efficient in reaching the target compared to DQN and A2C, which are more affected by complex environmental dynamics.

Keywords—navigation, 3D soft growing robot, PPO, DQN, A2C, reward function, observation space, action space.