

Penelitian ini membahas penerapan *Reinforcement Learning* (RL) untuk otomatisasi tugas manipulasi presisi *needle pick* pada platform simulasi robot bedah SurRoL (*Surgical Robot Learning*). Fokus utama penelitian diarahkan pada evaluasi algoritma *Proximal Policy Optimization* (PPO) sebagai metode *on-policy reinforcement learning* yang diimplementasikan secara langsung (*bare-metal*) tanpa modifikasi algoritmik tambahan, guna memahami karakteristik pembelajaran agen pada lingkungan simulasi bedah yang sensitif terhadap desain *reward function*.

Metodologi penelitian dilakukan sepenuhnya pada lingkungan simulasi SurRoL dengan menggunakan tiga variasi desain *reward function*, yaitu *sparse reward*, *less-sparse reward*, dan *staged reward*. Ketiga variasi tersebut merepresentasikan tingkat kepadatan sinyal pembelajaran yang berbeda, sehingga memungkinkan analisis pengaruh desain reward terhadap efisiensi pelatihan, dinamika pembaruan kebijakan *actor-critic*, serta tingkat keberhasilan agen dalam menyelesaikan tugas *needle pick*. Evaluasi performa dilakukan melalui analisis metrik internal PPO selama tahap pelatihan, serta metrik keberhasilan tugas pada tahap *test run*, yaitu *contact consistency* dan *average distance to goal*.

Hasil penelitian menunjukkan bahwa PPO dapat dilatih secara stabil pada lingkungan SurRoL dan mampu mempelajari perilaku manipulasi dasar tanpa memerlukan perubahan struktur algoritma. Desain *reward function* terbukti menjadi faktor dominan yang memengaruhi kualitas pembelajaran agen. Varian *less-sparse reward* menghasilkan dinamika pembelajaran yang lebih seimbang antara stabilitas dan otonomi agen, sedangkan *staged reward* menunjukkan indikasi hubungan asinkron antara pembaruan *actor* dan *critic* akibat perubahan distribusi reward secara bertahap, yang berdampak pada lambatnya konvergensi dan rendahnya performa pada tahap pengujian.

Kata kunci: Reinforcement Learning, PPO, SurRoL, Reward Shaping, Needle Pick.

This study investigates the application of Reinforcement Learning (RL) for automating a precision manipulation task, namely needle picking, within the Surgical Robot Learning (SurRoL) simulation platform. The research focuses on evaluating Proximal Policy Optimization (PPO) as an on-policy reinforcement learning algorithm implemented in a bare-metal manner, without additional algorithmic modifications, to analyze agent learning behavior in a surgical simulation environment that is highly sensitive to reward design.

The experiments are conducted entirely in simulation using three different reward function designs: sparse reward, less-sparse reward, and staged reward. These variations represent different levels of reward signal density, enabling a systematic analysis of their effects on training efficiency, actor-critic learning dynamics, and task success performance. Agent performance is evaluated through internal PPO training metrics and task-level test run metrics, including contact consistency and average distance to goal.

The results demonstrate that PPO can be trained stably in the SurRoL environment and is capable of learning fundamental manipulation behaviors without requiring algorithmic restructuring. Reward function design is shown to be the dominant factor influencing learning quality. The less-sparse reward variant provides a more balanced trade-off between learning stability and agent autonomy, while the staged reward approach exhibits asynchronous behavior between actor and critic updates due to non-stationary reward distributions, leading to slower convergence and reduced task success during testing.

Keywords : Reinforcement Learning, PPO, SurRoL, Reward Shaping, Needle Pick.