

Supplementary Material:

Contact-Guided Exploration for Non-Prehensile Locomanipulation with Multi-Critic RL

Simone Tolomei, Mayank Mittal, Franco Angelini, Manolo Garabini, Paolo Salaris, Marco Hutter

1 Individual Critic Heads Returns

Fig. 1 and Fig. 2 illustrate the individual training metrics for each of the three critic heads (task, exploration, and regularisation), showing the mean and standard deviation across 5 random seeds.

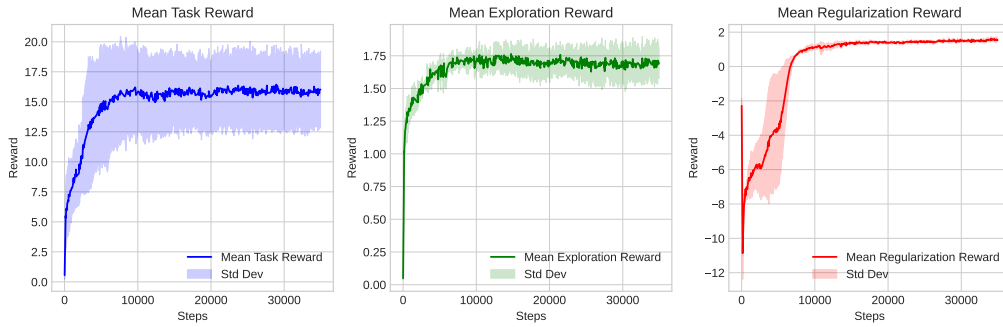


Figure 1: Return over steps, divided by groups.

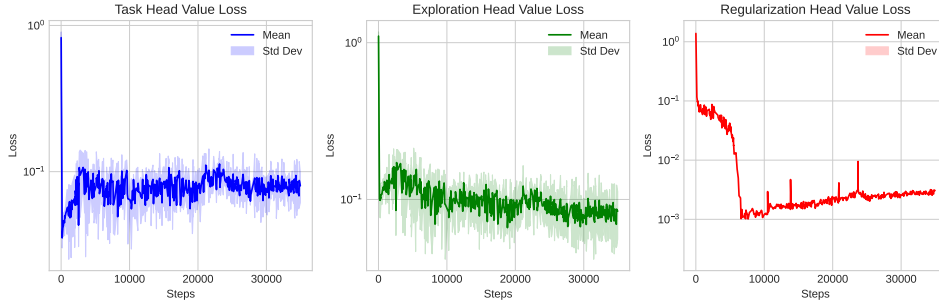


Figure 2: Value Function losses over steps, divided by critic head.

The effect of the weight schedule is clearly visible between 5k and 10k steps. As the regularisation weight increases, the policy adapts and increases the return of the regularisation head.