

Hyperparameter	Value
Learning Rate, $\alpha$	0.001
Learning Starts	100 Steps
Batch Size	100 Transitions
Tau, $\tau$	0.005
Gamma, $\gamma$	0.99
Training Frequency	1:Episode
Gradient Steps	$\propto$ Training Frequency
Action Noise, $\epsilon$	None
Policy Delay	1 : 2 Q-Function Updates
Target Policy Noise, $\epsilon$	0.2
Target Policy Clip, $c$	0.5
Seed	100 Random Seeds