Train policy π_ϕ
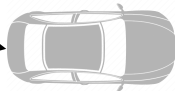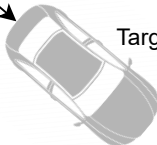
Target

Target

Target

Robot