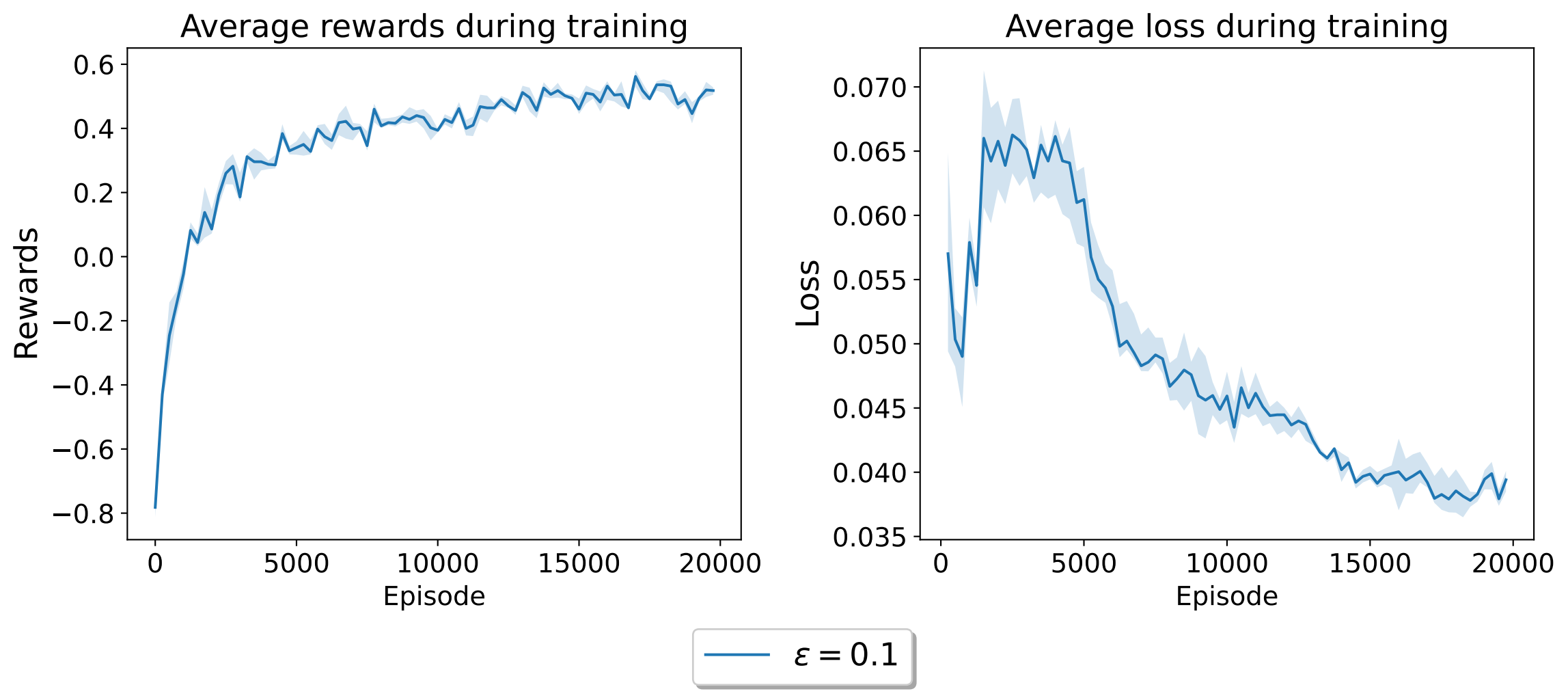
Average rewards during training
Rewards
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0
5000
10000
15000
20000
Episode
Average loss during training
Loss
0.070
0.065
0.060
0.055
0.050
0.045
0.040
0.035
0
5000
10000
15000
20000
Episode
$\varepsilon = 0.1$