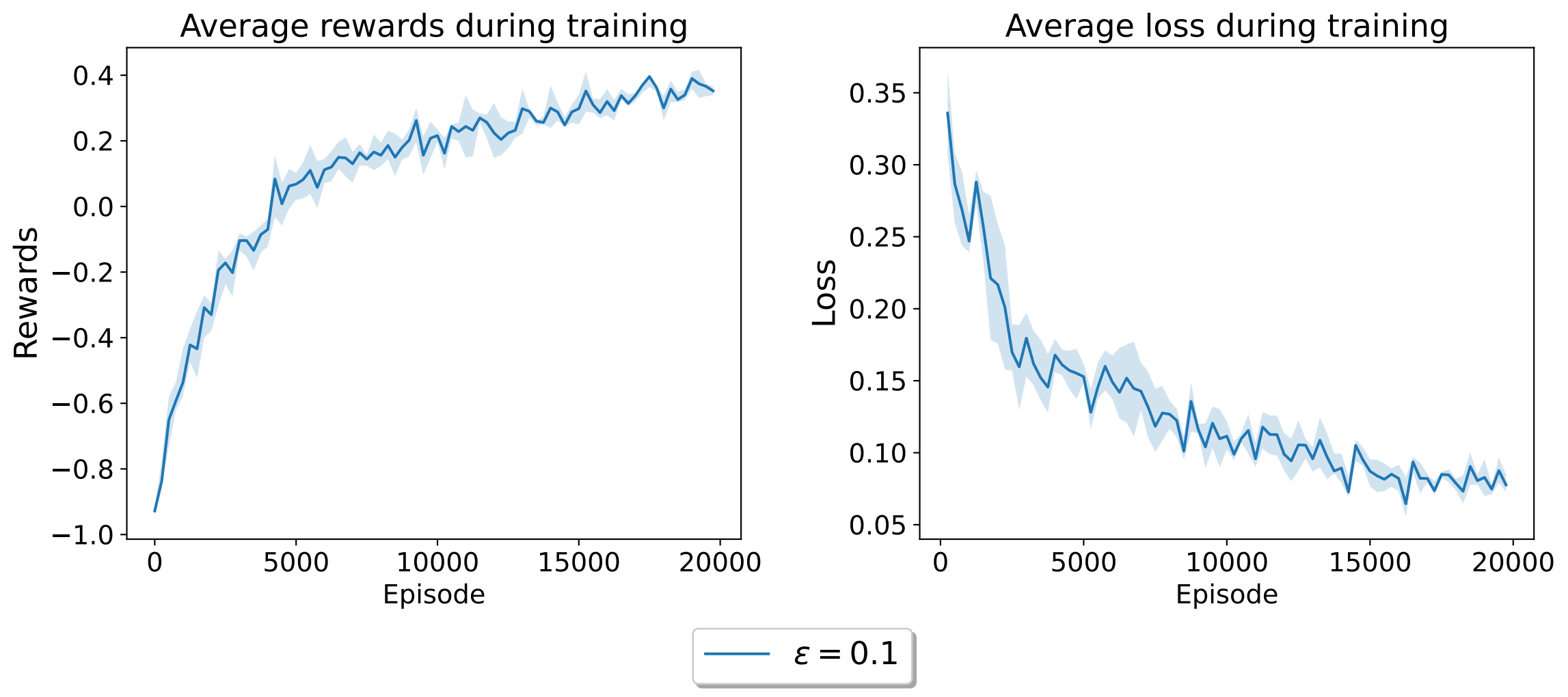
Average rewards during training
Rewards
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
0
5000
10000
15000
20000
Episode
Average loss during training
Loss
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
5000
10000
15000
20000
Episode
$\varepsilon = 0.1$